# Muestreo por importancia

Kevin Garcia - Alejandro Vargas - Alejandro Soto

2 de mayo de 2018

# Contenido

- Introducción
- Teoría
- Algoritmo
- Ejemplos

# Introducción

Aunque hay varios métodos para simular muestras de varias distribuciones, generalmente no es posible obtener una muestra i.i.d. directamente de la distribución a posteriori $h(\theta|x)$ y así hay necesidad de encontrar estrategias alternativas. Por ejemplo, una de esas estrategias posibles es la de simular de una distribución 'semejante' a la distribución a posteriori, para ello surge el muestreo por importancia que se tratará en esta presentación.

# Teoría

Sea $p(\theta)$ una función de densidad de la cual es fácil simular valores y que aproxime $h(\theta|x) = cf(x|\theta)h(\theta)$. Entonces

$$\int g(\theta)h(\theta|x)d\theta = \frac{\int g(\theta)f(x|\theta)h(\theta)d\theta}{\int f(x|\theta)h(\theta)d\theta}$$

$$\int g(\theta)h(\theta|x)d\theta = \frac{\int g(\theta)\frac{f(x|\theta)h(\theta)}{p(\theta)}p(\theta)d\theta}{\int \frac{f(x|\theta)h(\theta)}{p(\theta)}p(\theta)d\theta}$$

$$\int g(\theta)h(\theta|x)d\theta = \frac{\int g(\theta)\omega(\theta)p(\theta)d\theta}{\int \omega(\theta)p(\theta)d\theta}$$

# Teoría

Si se obtiene una muestra $\theta_1, \theta_2, ..., \theta_n$ de $p(\theta)$, se puede aplicar el método de Monte Carlo, obteniéndose entonces como aproximación de $E[g(\theta)|x]$

$$\hat{E}[g(\theta)|x] = \frac{1}{\sum_{i=1}^{n} \omega_1} \sum_{i=1}^{n} \omega_i g(\theta_i)$$

donde $\omega_i = f(x|\theta_i)h(\theta_i)/p(\theta_i)$

El método de muestreo por importancia atribuye asi mas peso a regiones donde $p(\theta) < h(\theta|x)$ y menos peso a regiones donde $p(\theta) > h(\theta|x)$. Geweke(1989) muestra que si el soporte de $p(\theta)$ incluye el soporte de $h(\theta|x)$, los $\theta_i$ son una muestra i.i.d. de $p(\theta)$, y $\int g(\theta)h(\theta|x)d\theta$ existe y es finita, entonces

# Teoría

$$\frac{1}{\sum_{i=1}^{n} \omega_i} \sum_{i=1}^{n} \omega_i g(\theta_i) \to \int g(\theta) h(\theta|x) d\theta$$

Con un error estándar de Monte Carlo estimado por:

$$\frac{1}{\sum_{j=1}^{n} \omega_j} \left[ \sum_{i=1}^{n} \left\{ g(\theta_i) - \frac{1}{\sum_{j=1}^{n} \omega_j} \sum_{i=1}^{n} \omega_i g(\theta_i) \right\}^2 \omega_i^2 \right]^{1/2}$$

La razón de convergencia depende de cuán bien $p(\theta)$, la función de importancia, imita $h(\theta|x)$. "Buenas" propiedades de la función de importancia son:

1. Simplicidad en la generación de números pseudo-aleatorios
2. Tener colas más pesadas que $h(\cdot|x)$
3. Ser una buena aproximación a $h(\cdot|x)$

# Teoría

Se debe señalar que para aplicar esta metodologia sólo hay necesidad de exigir que $h(\theta|x)$ sea conocida a menos de la constante de proporcionalidad, es decir, basta considerar $f(\theta|x)h(\theta)$.Esta observación, también aplicable a la función de importancia, es importante ya que evita la necesidad de calcular la integral necesaria para la obtención de la respectiva constante de proporcionalidad.

# Algoritmo

1. Simular $\theta_1, \theta_2, ..., \theta_m \sim iid\ p(\theta)$
2. Se calcula $\omega_i = \frac{h(\theta_i|y)}{p(\theta_i)}$
3. Se calcula $\frac{1}{\sum_{i=1}^{m}\omega_i}\sum_{i=1}^{m}\omega_i g(\theta_i)$, con

- $g(\theta) = \theta$ para el cálculo aproximado del valor medio de la distribución a posteriori
- $g(\theta) = \theta^2$ para obtener una aproximación de $E(\theta^2)$ de la distribución a posteriori.
  Entonces, $V(\theta) = E(\theta^2) - E^2(\theta)$

# Ejemplo 1

Sea $X_i \sim N(x_i|\theta, 1)$. La función de verosimilitud es una distribución $N(\bar{x}|\theta, \frac{1}{n})$. Y tomando una distribución a priori $\pi(\theta) \propto 1$. Se quiere estimar $E(\theta|x, \sigma^2)$.

Para este ejemplo se utilizo una función de importancia $p(\theta) \sim t_n$
Entonces tenemos que $\theta|x, \sigma^2 \sim N(\bar{x}, \frac{1}{n})$
Para tener el $\bar{x}$ y $\sigma^2$ generamos 10000 valores de x con distribución N(0,1), los resultados fueron:

$$\bar{x} = 0{,}007557463, \sigma^2 = \frac{1}{10000} = 0{,}0001$$

# Ejemplo 1

Generamos 10000 $\theta_i$ con distribución $p(\theta) \sim t_n$

Posteriormente se calculo $\omega_i = \frac{f(\theta_i|x,\sigma^2)}{p(\theta_i)}$

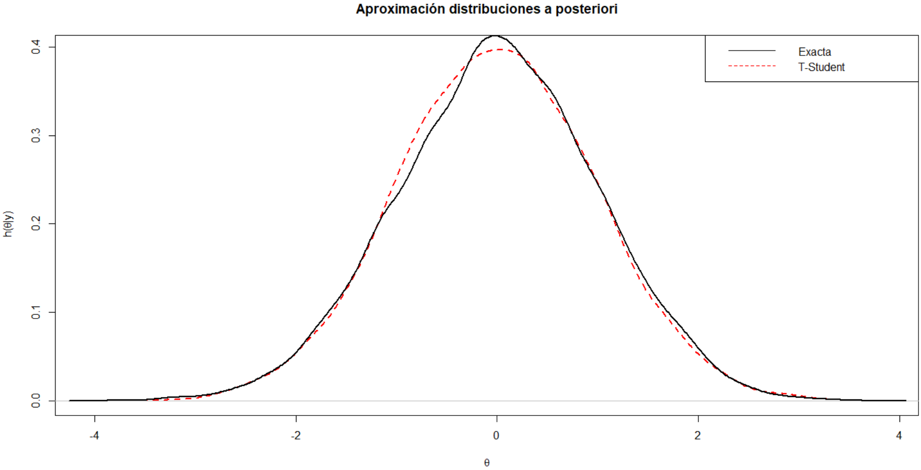
Finalmente para estimar la esperanza a posteriori, se calculo

$$\frac{1}{\sum\limits_{i=1}^{N} \omega_i} \sum_{i=1}^{N} \omega_i \theta_i = 0{,}008330487$$

Y, para estimar la varianza a posteriori, se calculo

$$\frac{1}{\sum\limits_{i=1}^{N} \omega_i} \sum_{i=1}^{N} \omega_i \theta_i^2 = 0{,}0001650637$$

# Ejemplo 1



Aproximación distribuciones a posteriori
Exacta
T-Student
h(θ|y)
0.0
0.1
0.2
0.3
0.4
-4
-2
0
2
4
θ

# Ejemplo 1

Comparando los resultados, tenemos:

| | $E(\theta \mid x, \sigma^2)$ | $V(\theta \mid x, \sigma^2)$ |
|---|---|---|
| Teorica | 0.007557463 | 0.0001 |
| Función de importancia $t_n$ | 0.008330487 | 0.0001650637 |

# Ejemplo 2

Usando una función de importancia uniforme aproximar la media y la varianza de la distribución a posteriori de $\theta$, para la funcion a posteriori $f(\theta|x) \propto \theta^9(1-\theta)^3, 0 < \theta < 1$. Para obtener la media y la varianza a posteriori:

Generamos 10000 $\theta_i$ con distribución $p(\theta) \sim U(0,1)$

Posteriormente se calculo $\omega_i = \frac{f(\theta_i|x,\sigma^2)}{p(\theta_i)}$

Finalmente para estimar la esperanza a posteriori, se calculo

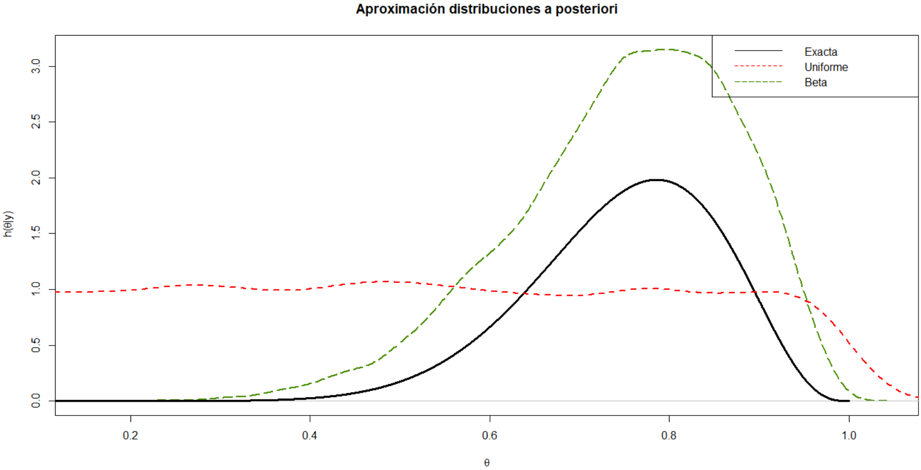
$$\frac{1}{\sum_{i=1}^{N} \omega_i} \sum_{i=1}^{N} \omega_i \theta_i = 0{,}7324333$$

Y, para estimar la varianza a posteriori, se calculo

$$\frac{1}{\sum_{i=1}^{N} \omega_i} \sum_{i=1}^{N} \omega_i \theta_i^2 = 0{,}5489597$$

Entonces, $V(\theta|x,\sigma^2) = 0{,}01250114$

# Ejemplo 2



Aproximación distribuciones a posteriori
Exacta
Uniforme
Beta
h(θ|y)
θ
0.0
0.5
1.0
1.5
2.0
2.5
3.0
0.2
0.4
0.6
0.8
1.0

# Ejemplo 2

Comparando los resultados, tenemos:

| | $E(\theta\|x, \sigma^2)$ | $V(\theta\|x, \sigma^2)$ |
|---|---|---|
| Teorica | 0.7142857 | 0.01360544 |
| Función de importancia Uniforme | 0.7148448 | 0.01362596 |
| Función de importancia Beta | 0.7142896 | 0.01362596 |

# Ejemplo 3

Segun un modelo genético, los animales de una determinada especie están distribuidos en 4 categorías, de acuerdo a las probabilidades.

$$p_1 = \frac{2+\theta}{4}, p_2 = \frac{1-\theta}{4}, p_3 = \frac{1-\theta}{4}, p_4 = \frac{\theta}{4},$$

donde $0 \leq \theta \leq 1$ es un parámetro desconocido, sobre el cuál se quieren hacer inferencias. Asuma que se adopta para $\theta$ una distribución a priori $Be(a, b)$ y que para una muestra de tamaño N se observaran $y_i$ animales en la i-ésima categoría $(i = 1, ..., 4, \sum_i y_i = N)$. En estas condiciones la distribución a posteriori para $\theta$ es:

$$h(\theta|y) \propto (2+\theta)^{y_1}(1-\theta)^{y_2+y_3+b-1}\theta^{y_4+a-1}, 0 \leq \theta \leq 1$$

# Ejemplo 3

Para obtener los parámetros de las funciones de importancia a simular, se debe solucionar la ecuación $L'(\theta) = 0$ para obtener el valor medio $\hat{\theta}$ y para la varianza $\sigma^2 = \left\{-L''(\hat{\theta})\right\}^{-1}$.
De la función a posteriori tenemos:

$$L(\theta|y) = logh(\theta|y) \propto y_1 log(2+\theta) + (y_2 + y_3 + b - 1) log(1-\theta)$$

$$+(y_4 + a - 1) log(\theta)$$

$$L'(\theta) = \frac{y_1}{2+\theta} - \frac{y_2 + y_3 + b - 1}{1-\theta} + \frac{y_4 + a - 1}{\theta}$$

$$L''(\theta) = \frac{y_1}{(2+\theta)^2} - \frac{y_2 + y_3 + b - 1}{(1-\theta)^2} + \frac{y_4 + a - 1}{\theta^2}$$

# Ejemplo 3

Utilizando una distribución a priori $Be(a = 1, b = 1)$ y considerando dos muestras de tamaños diferentes para las cuales se obtienen los siguientes datos por categorias, N=197, y=(125,18,20,34) y N=20, y=(14,0,1,5).
Veamos como hacer uso del concepto de función de importancia para obtener la distribución a posteriori y calcular el valor medio y varianza a posteriori.

# Ejemplo 3

- N=197:

$$L'(\theta) = 0 \rightarrow \hat{\theta} = 0{,}6268$$

$$\hat{\sigma^2} = \{-L''(0{,}6268)\}^{-1} = 0{,}002649$$

Entonces, las funciones de importancia son, $N(0{,}6268, 0{,}002649)$ y $Beta(54{,}723205, 32{,}58249)$

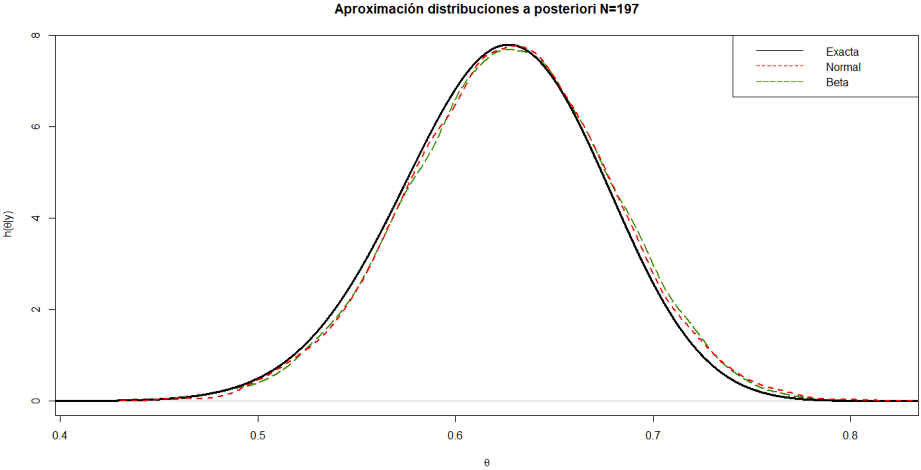
- N=20:

$$L'(\theta) = 0 \rightarrow \hat{\theta} = 0{,}9034$$

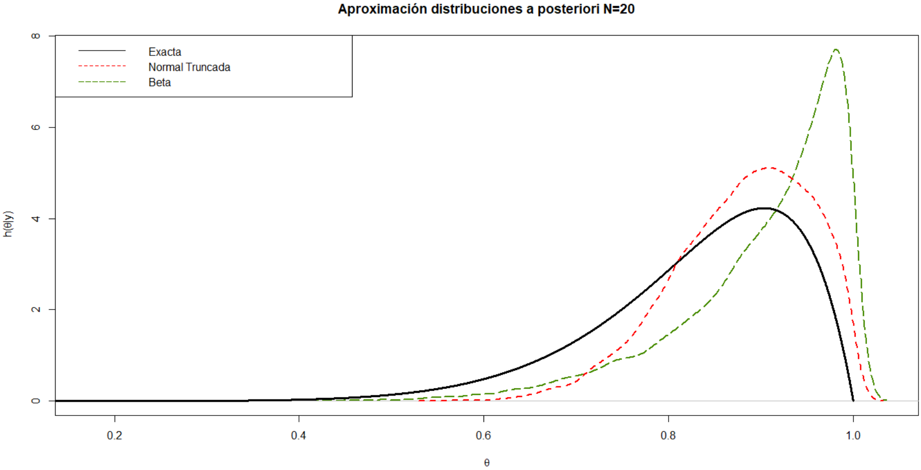
$$\hat{\sigma^2} = \{-L''(0{,}6268)\}^{-1} = 0{,}008693$$

Entonces, las funciones de importancia son, $N(0{,}9034, 0{,}008693)I_{[-\infty,1]}$ y $Beta(8{,}165772, 0{,}873161)$

# Ejemplo 3


Aproximación distribuciones a posteriori N=197
Exacta
Normal
Beta
h(θ|y)
θ
0
2
4
6
8
0.4
0.5
0.6
0.7
0.8

# Ejemplo 3



Aproximación distribuciones a posteriori N=20
Exacta
Normal Truncada
Beta
h(θ|y)
θ
0
2
4
6
8
0.2
0.4
0.6
0.8
1.0

# Ejemplo 3

| | N=197 | | N=20 | |
|---|---|---|---|---|
| Función de importancia | $E(\theta\|y)$ | $V(\theta\|y)$ | $E(\theta\|y)$ | $V(\theta\|y)$ |
| Beta | 0.6230844 | 0.002572936 | 0.8306012 | 0.01188336 |
| Normal | 0.6232495 | 0.00257044 | 0.8377618 | 0.009922389 |
| Exacta | 0.6228062 | 0.002594837 | 0.8311239 | 0.01165126 |

Como se puede observar, en el caso de la primera muestra, el método de muestreo por importancia proporciona buenas aproximaciones, para ambas funciones de importancia consideradas. El mismo ya no ocurre en el caso de la segunda muestra, donde la función Normal (incluso truncada) no es adecuada. Sin embargo, la función Beta proporciona una buena aproximación.

# Códigos R

## Ejemplo 1

```
#Ejemplo 1:
N=10000
X<-rnorm(N)
xbarra=mean(X)
varianza=1/N
thetai<-rt(N,N)
w<-dnorm(thetai,xbarra,sqrt(varianza))/dt(thetai,N)
E<-(1/sum(w))*(sum(w*thetai))
V<-(1/sum(w))*(sum(w*(thetai^2)))
```

# Códigos R

## Ejemplo 2

```
#Ejemplo 2:
funcion<-function(x){(x^9)*((1-x)^3)}
tetai<-runif(N)
w1<-funcion(tetai)/dunif(tetai)
E1<-(1/sum(w1))*(sum(w1*tetai))  #Esperanza estimada
V1<-(1/sum(w1))*(sum(w1*(tetai^2))) #Varianza estimada
```