

Muestreo por importancia

Kevin Garcia - Alejandro Vargas - Alejandro Soto

4 de mayo de 2018

Contenido

- Introducción
- Muestreo por importancia
- Teoría
- Algoritmo
- Ejemplos
- Bibliografía
- Anexos (Códigos R)

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.

Muestreo por importancia

El muestreo de importancia es más que solo un método de reducción de varianza. Puede ser utilizado para estudiar una distribución mientras se toma el muestreo de otra. Como resultado, puede utilizar el muestreo de importancia como una alternativa al muestreo de rechazo de aceptación, como un método para el análisis de sensibilidad y como la base para algunos métodos de cálculo de constantes de normalización de las densidades de probabilidad. Muestreo de importancia también es un prerequisite importante para Monte Carlo secuencial.

Muestreo por importancia

La idea detrás del muestreo de importancia es que ciertos valores de las variables aleatorias de entrada en una simulación tienen más impacto en el parámetro que se estima que otros. Si estos valores 'importantes' son enfatizados por muestreo con mayor frecuencia, entonces la varianza del estimador puede ser reducida. Por lo tanto, la metodología básica en el muestreo de importancia es elegir una distribución que 'aliente' los valores importantes.

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

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

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

El método de muestreo por importancia atribuye así 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 θ_i son una muestra i.i.d. de $p(\theta)$, y $\int g(\theta)h(\theta|x)d\theta$ existe y es finita, entonces

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

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 $h(\cdot|x)$

Se debe señalar que para aplicar esta metodología 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 $\theta_1, \theta_2, \dots, \theta_m \sim iid p(\theta)$
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 σ^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 θ_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

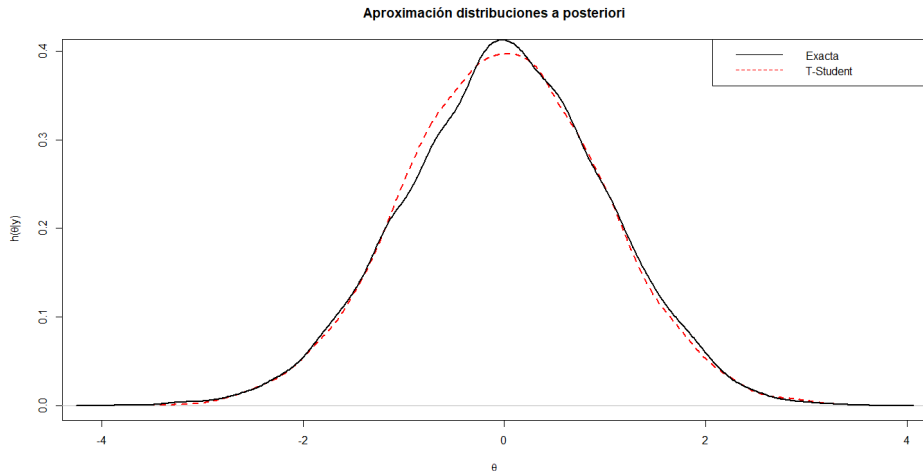
$$\frac{1}{\sum_{i=1}^N \omega_i} \sum_{i=1}^N \omega_i \theta_i = 0,01375921$$

Y, para estimar la varianza a posteriori, se calculo

$$\frac{1}{\sum_{i=1}^N \omega_i} \sum_{i=1}^N \omega_i \theta_i^2 = 0,0002915643$$

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

Ejemplo 1



Ejemplo 1

Comparando los resultados, tenemos:

	$E(\theta x, \sigma^2)$	$V(\theta x, \sigma^2)$
Teórica	0.01364327	0.0001
Función de importancia t_n	0.01375921	0.0001022485

Ejemplo 2

Usando una función de importancia uniforme aproximar la media y la varianza de la distribución a posteriori de θ , para la función 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 θ_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

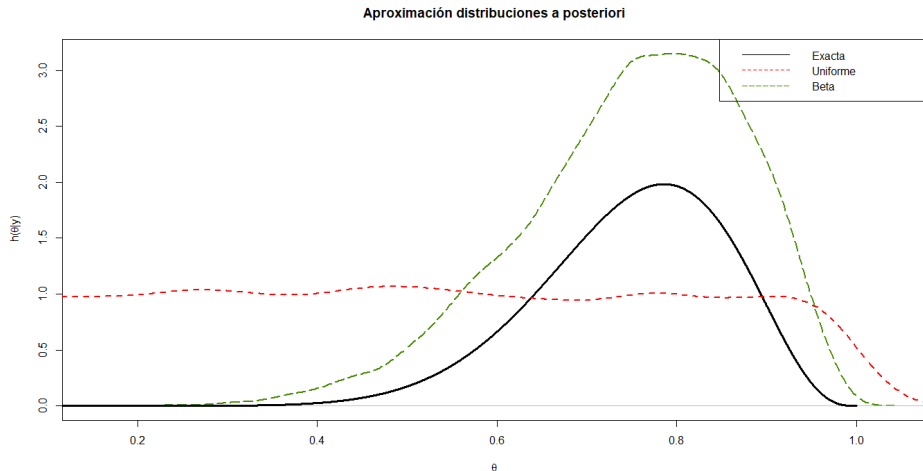
$$\frac{1}{\sum_{i=1}^N \omega_i} \sum_{i=1}^N \omega_i \theta_i = 0,7324333$$

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



Ejemplo 2

Comparando los resultados, tenemos:

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

Ejemplo 3

Según 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 θ 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 ($i = 1, \dots, 4, \sum_i y_i = N$). En estas condiciones la distribución a posteriori para θ 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:

$$L(\theta|y) = \log h(\theta|y) \propto y_1 \log(2 + \theta) + (y_2 + y_3 + b - 1) \log(1 - \theta) \\ + (y_4 + a - 1) \log(\theta)$$

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

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

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 categorías, $N=197$, $y=(125,18,20,34)$ 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 $Beta(54,723205, 32,58249)$

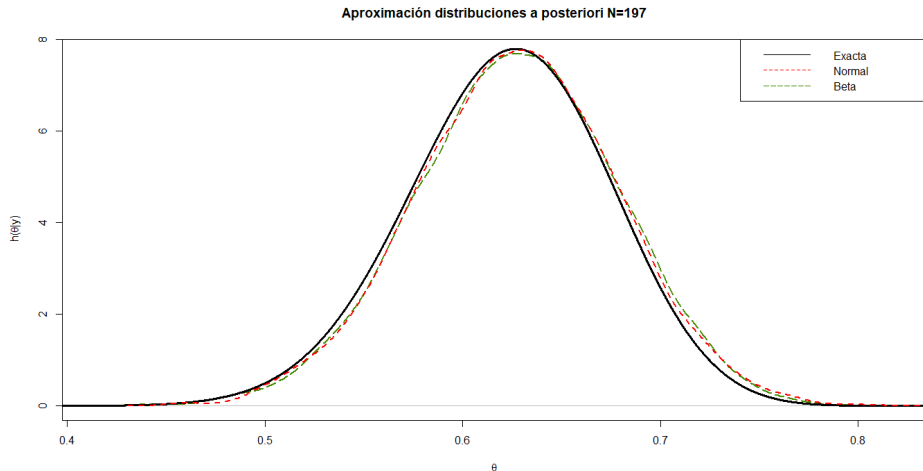
- N=20:

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

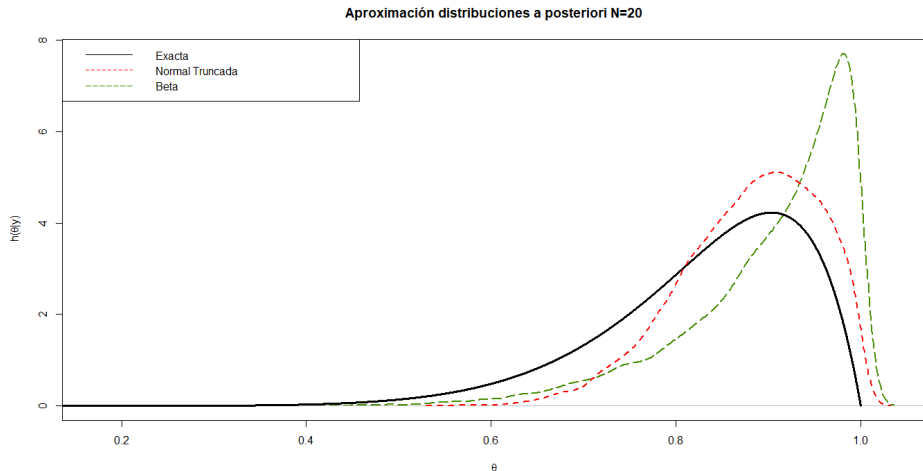
$$\hat{\sigma}^2 = \{-L''(0,9034)\}^{-1} = 0,008693$$

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

Ejemplo 3



Ejemplo 3



Ejemplo 3

Función de importancia	N=197		N=20	
	$E(\theta y)$	$V(\theta y)$	$E(\theta y)$	$V(\theta y)$
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

Ejemplo 3

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.

Bibliografía

1. ESTATÍSTICA BAYESIANA, Paulino, Turkman y Murteira. 2003.
2. Importance Sampling Applications in Communications and Detection, Rajan Srinivasan. 2002.
3. Monte Carlo Methods and Importance Sampling, Eric C. Anderson. 1999.
4. Tesis: MUESTREO POR IMPORTANCIA Y REDUCCIÓN RECURSIVA DE VARIANZA PARA LA EVALUACIÓN DE LA CONFIABILIDAD DIÁMETRO ACOTADA, Fernanda González. 2013.
5. Importance sampling. Wikipedia.

Códigos R

Ejemplo 1

#Ejemplo 1:

N=10000

X<-rnorm(N) #Esto se hizo para tener la media y la varian

xbarra=mean(X) #Media teorica

varianza=1/N #Varianza teorica

theta1<-rt(N,N) #Tetas generados a partir de la función

w<-dnorm(theta1,xbarra,sqrt(varianza))/dt(theta1,N) #Peso

E<-(1/sum(w))*(sum(w*theta1)) #Esperanza estimada 0.013

E2<-(1/sum(w))*(sum(w*(theta1^2))) #Esperanza^2 estimada

V<-E2-(E^2) #Varianza 0.0001022485

Códigos R

Ejemplo 2

```
funcion<-function(x){(x^9)*((1-x)^3)}  
funcion2exacta<-function(x){(gamma(14)/(gamma(10)*gamma(4)  
integrate(funcion,0,1) #Constante de integración  
integrate(funcion2exacta,0,1) #Integral de la densidad con  
Eteorica<-integrate(funcion2exacta,0,1) #0.7142857  
E2teorica<-integrate(funcion2exacta,0,1) #0.5238095  
Vteorica<-0.5238095-(0.7142857^2) #Resultado: 0.01360544  
  
teta2<-runif(N) #Funcion de importancia U(0,1)  
w1<-funcion(teta2)/dunif(teta2,0,1)  
E1<-(1/sum(w1))*(sum(w1*teta2)) #Esperanza estimada 0.7  
E21<-(1/sum(w1))*(sum(w1*(teta2^2))) #0.5247885  
V1<-E21-(E1^2) #Varianza estimada 0.01378547
```

Códigos R

Ejemplo 2

```
teta22<-rbeta(N,8.25,2.75) #Función de importancia Beta(8
w12<-funcion(teta22)/dbeta(teta22,8.25,2.75)
E12<-(1/sum(w12))*(sum(w12*teta22)) #Esperanza estimada
E212<-(1/sum(w12))*(sum(w12*(teta22^2))) #Varianza estimada
V12<-E212-(E12^2) #0.01362596
```

Códigos R

Ejemplo 3

```
#N=197
m=10000
N=197
y1=125,y2=18,y3=20,y4=34
a=b=1
h<-function(x){((2+x)^y1)*((1-x)^(y2+y3+b-1))*x^(y4+a-1)}
hexacta<-function(x){(1/2.357695e+28)*((2+x)^y1)*((1-x)^(y2+y3+b-1))*x^(y4+a-1)}
#exacta
integrate(h,0,1)
Esp teor<-integrate(hexacta,0,1) #0.6228062
Esp2teor<-integrate(hexacta,0,1) #0.3904824
Varteori<-0.3904824-(0.6228062)^2 #0.002594837
```

Códigos R

Ejemplo 3

```
#Función de importancia Beta
p<-rbeta(m,54.723205,32.58249)
w<-h(p)/dbeta(p,54.723205,32.58249)
E<-(1/sum(w))*(sum(w*p)) #0.6230844
EB2<-(1/sum(w))*(sum(w*(p^2))) #0.3908071 #E(x^2)
VB<-EB2-(E^2) #0.00257293

#Función de importancia normal
p1<-rnorm(m,0.6268,sqrt(0.002649))
w1<-h(p1)/dnorm(p1,0.6268,sqrt(0.002649))
E1<-(1/sum(w1))*(sum(w1*p1)) #0.623249
EN2<-(1/sum(w1))*(sum(w1*(p1^2))) #0.3910104 #E(x^2)
VN<-EN2-(E1^2) #0.0025704
```


Códigos R

Ejemplo 3

```
#N=20
y11=14
y22=0
y33=1
y44=5
h1<-function(x){((2+x)^y11)*((1-x)^(y22+y33+b-1))*x^(y44+
h1exacta<-function(x){(1/41575.13)*((2+x)^y11)*((1-x)^(y2
#exacta
integrate(h1,0,1) #Constante de integración
Esp2teor00<-integrate(h1exacta,0,1) #0.8311239
Esp2teor00<-integrate(h1exacta,0,1) #0.7024182
Varteori00<-0.7024182-(0.8311239)^2 #0.01165126
```

Códigos R

Ejemplo 3

```
#Función de importancia Beta
p00<-rbeta(m,8.165772,0.873161)
w00<-h1(p00)/dbeta(p00,8.165772,0.873161)
E00<-(1/sum(w00))*(sum(w00*p00)) #0.8306012
E200<-(1/sum(w00))*(sum(w00*p00^2)) #0.7017816
V00<-E200-(E00^2) #0.01188336

#Función de importancia normal Truncada:
p111<-rtruncnorm(m,a=-Inf,b=1,0.9034,sqrt(0.008693))
w111<-h1(p111)/dtruncnorm(p111,a=-Inf,b=1,0.9034,sqrt(0.008693))
E111<-(1/sum(w111))*(sum(w111*p111)) #0.8377618
E2111<-(1/sum(w111))*(sum(w111*p111^2)) #0.7117673
V111<-E2111-(E111^2) #0.009922389
```