

Muestreo por importancia

Kevin Garcia - Alejandro Vargas - Alejandro Soto

13 de abril de 2018

Contenido

-
-
-

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.

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

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

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

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

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 más 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: (i) simplicidad en la generación de números pseudo-aleatorios; (ii) tener colas más pesadas que $h(\cdot|x)$; (iii) 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.

Ejemplo

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. Supongamos que se tiene una función de densidad a posteriori de la siguiente forma:

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

y

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

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

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

Una función de importancia bastante utilizada es la función de densidad Normal, ya que lo que se pretende es simular de una distribución que aproxime la distribución a posteriori. Sin embargo hay situaciones en que la aproximación por la normal no es adecuada y por lo tanto otra función de importancia debe ser considerada. Una representación gráfica de la verosimilitud puede ayudar en la selección; en este caso, dado que el parámetro para el que se pretende obtener la distribución a posteriori varía en el intervalo $[0, 1]$, la función de densidad Beta también puede considerarse adecuada como candidata a función de importancia.

Se procederá entonces al método de muestreo por importancia utilizando las dos funciones, Normal y Beta, para las dos muestras en estudio. Para ello es necesario simular muestras de esas distribuciones. Se designa por $p(\theta)$ la función de importancia en consideración.

Como se dijo anteriormente, sea $\hat{\theta}$ el valor de θ para el cual $L'(\theta) = 0$, y $\hat{\sigma}^2 = \left\{-L''(\hat{\theta})\right\}^{-1}$. Se consideran estos valores como aproximaciones, respectivamente, para el valor medio y la varianza de la distribución a posteriori. Estos valores serán necesarios para obtener los parámetros de las distribuciones a simular. Se procede entonces de la siguiente manera:

1. Se simula $\theta_1, \theta_2, \dots, \theta_m \sim iid p(\theta)$
2. Se calcula $\omega_i = \frac{h(\theta_i|y)}{p(\theta_i)}$
3. Se calcula $\frac{1}{\sum_{i=1}^m} \sum_{i=1}^m \omega_i g(\theta_i)$, 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 la varianza de la distribución a posteriori.

Se observa que en el procedimiento anterior basta conocer el núcleo de la distribución posterior, es decir, basta conocer $h(\theta|y)$ a menos de la constante de proporcionalidad. También se puede obtener una mejor aproximación a la función de densidad a posteriori, asignando masa $\omega_i / \sum_{j=1}^m \omega_j$ a los valores simulados θ_i .

INSERTAR IMÁGENES DE LOS RESULTADOS

En las figuras anteriores se encuentran los gráficos de la distribución a posteriori exacta obtenida usando métodos de integración, así como las aproximaciones obtenidas usando método de muestreo por importancia, considerando la función de densidad Normal y la función de densidad Beta, para las dos muestras en estudio. Se utilizó $m=10000$ y una distribución a priori uniforme. En el caso de la segunda muestra, dado que eran simulados valores de θ mayores que 1, se usó una distribución normal truncada en 1.

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.