

# Muestreo por importancia

Kevin Garcia - Alejandro Vargas - Alejandro Soto

3 de abril de 2018

# Contenido

- 
- 
-

La simulación rápida con muestreo por importancia (importance sampling) IS es esencialmente un procedimiento forzoso de Monte Carlo diseñado para acelerar la ocurrencia de eventos raros. El desarrollo de este método de simulación de análisis de fenómenos científicos generalmente se atribuye al matemático von Neumann y otros. Desde su inicio, la simulación MC ha encontrado una amplia gama de empleo, desde la termodinámica estadística en sistemas desordenados hasta el análisis y diseño de estructuras de ingeniería caracterizado por alta complejidad. De hecho, cada vez que un problema de ingeniería es analíticamente intratable (que a menudo es el caso) y una solución mediante técnicas numéricas prohibitivamente costosa computacionalmente, un último recurso para determinar las características de entrada-salida de, o estados dentro de, un sistema es llevar a cabo una simulación.

Este no es un método para generar muestras. Este es un método para calcular la esperanza de  $h(\theta)$ . Asumamos que nuestra densidad unidimensional objetivo es  $p(\theta)$ , y de la cual tenemos su kernel, digamos  $p^*(\theta)$  tal que

$$p(\theta) = \frac{p^*(\theta)}{Z}$$

Donde  $Z$  es una constante de normalización.

Supongamos que muestrear directamente de  $p(\theta)$  es muy complicado. Ahora asumamos que existe una distribución  $q(\theta)$  de la cual sabemos es fácil muestrear y que tiene el mismo soporte que  $p$ . La densidad  $q$  es llamada la densidad muestreadora.

En el muestreo por importancia procedemos de la siguiente manera:

- 1 Generamos  $R$  muestras  $\theta^{(1)}, \theta^{(2)}, \dots, \theta^{(R)}$  de  $q(\theta)$
- 2 Calculamos los pesos

$$w_r = \frac{p^*(\theta^{(r)})}{q(\theta^{(r)})}$$

- 3 Utilizamos los pesos anteriores para ajustar la “importancia” de cada punto en nuestro estimador así:

$$\hat{\phi} = \sum_{r=1}^R \frac{w_r}{\sum_{j=1}^R w_j} h(\theta^{(r)})$$

El muestreo por importancia es un método relacionado con el muestreo de rechazo, que se utiliza para calcular las esperanzas utilizando una muestra aleatoria extraída de una aproximación a la distribución objetivo.

Supongamos que estamos interesados en  $E(h(\theta) | y)$ , pero no podemos generar extracciones o muestras aleatorias de  $\theta$  a partir de  $p(\theta | y)$  y, por lo tanto, no podemos evaluar la integral mediante un promedio simple de valores simulados.

Si  $g(\theta)$  es una densidad de probabilidad a partir de la cual podemos generar muestras aleatorias, entonces podemos escribir.

$$E(h(\theta | y)) = \frac{\int h(\theta)q(\theta | y)d\theta}{\int q(\theta | y)d\theta} = \frac{\int \left[ \frac{h(\theta)q(\theta|y)}{g(\theta)} \right] g(\theta)d\theta}{\int \left[ \frac{q(\theta|y)}{g(\theta)} \right] g(\theta)d\theta}$$

Que puede ser estimado usando  $S$  muestras  $\theta^1, \dots, \theta^S$  de  $g(\theta)$  por la expresión,

$$\frac{\frac{1}{S} \sum_{s=1}^S h(\theta^s) w(\theta^s)}{\frac{1}{S} \sum_{s=1}^S w(\theta^s)}$$

donde el factor,

$$w(\theta^s) = \frac{q(\theta^s | y)}{g(\theta^s)}$$

Se llaman razones de importancia o ponderaciones de importancia. Recuerde que  $q$  es nuestra notación general para densidades no normalizadas; es decir,  $q(\theta | y)$  es igual a  $p(\theta | y)$  por algún factor que no depende en  $\theta$ .

En general, es aconsejable utilizar el mismo conjunto de muestras aleatoria tanto para el numerador como para el denominador de para reducir el error de muestreo en la estimación.