

Método Monte Carlo: determinación del número de iteraciones necesarias

Juan José Goyeneche

Instituto de Estadística
Facultad de Ciencias Económicas y de Administración
Universidad de la República

Noviembre de 2010

Resumen

El método de Monte Carlo se usa principalmente para estimar parámetros en modelos que involucran una o más variables aleatorias. En general estos modelos son muy complicados para resolver analíticamente.

Intuitivamente, cuantas más iteraciones se realicen, mejor será la precisión con que se pueden presentar los resultados. El objetivo de este artículo es presentar una forma de determinar R , el número de iteraciones necesarias.

Introducción

El método Monte Carlo consiste en simular el comportamiento de un modelo aleatorio a los efectos de aproximar la distribución en el muestreo de algunas estadísticas relacionadas con el mismo. La verdadera distribución en el muestreo de estas estadísticas es en general muy complicada de resolver analíticamente. Si bien el método está relacionado al uso de computadoras y la capacidad de éstas de generar números aleatorios, Daniel Peña menciona que el método ya fue usado por Student tomando 750 muestras de tamaño 4 de unas mediciones de dedos de 3000 delincuentes (Peña, (2001), pág. 40) a los efectos de estudiar la distribución t .

Número de Simulaciones y Precisión en los cálculos

Se simula el lanzamiento de dos dados, uno celeste y uno verde. Se consideran las variables aleatorias C_i y V_i , puntajes del dado celeste y del dado

verde en la i -ésima “tirada” respectivamente. Sean las variables aleatorias $A_i = I_{\{C_i \leq 5\}}$ y $B_i = I_{\{C_i \leq V_i\}}$. A lo largo de una serie de R tiradas los promedios $a_{(\cdot)} = R^{-1} \sum_{i=1}^{i=R} a_i$ y $b_{(\cdot)} = R^{-1} \sum_{i=1}^{i=R} b_i$ convergen en probabilidad a

$$\mu_A = E(A) = P(A = 1) = 5/6 \doteq 0,83333$$

y a

$$\mu_B = E(B) = P(B = 1) = (6 + 5 + 4 + 3 + 2 + 1)/36 = 21/36 \doteq 0,58333$$

respectivamente. En general los valores de $E(A)$ y $E(B)$ son desconocidos, así como los demás parámetros de la distribución de A y de B (y por eso es que hacemos simulación Monte Carlo).

La pregunta que pretendemos contestar es cual sería el valor de R necesario para aproximar los valores de μ_A y de μ_B con 3 o más cifras significativas con un 99 % de confianza, o en general, con una precisión ε y una confianza de $1 - \alpha$.

Esto es, $P(|a_{(\cdot)} - \mu_A| > \varepsilon) \leq \alpha$ y $P(|b_{(\cdot)} - \mu_B| > \varepsilon) \leq \alpha$. Determinaremos primero el valor de R necesario para asegurar $P(|a_{(\cdot)} - \mu_A| > \varepsilon) \leq \alpha$.

Si R es suficientemente grande podemos aproximar las distribuciones de $a_{(\cdot)}$ con la distribución normal. El número de repeticiones necesarias surge entonces de hacer

$$\varepsilon / \sqrt{Var(a_{(\cdot)})} > Z_{1-\alpha/2},$$

donde $Var(a_{(\cdot)})$ es la varianza del promedio muestral de las a_i y $Z_{1-\alpha/2}$ es el quantil $1 - \alpha/2$ de la distribución normal estandarizada. La $Var(a_{(\cdot)}) = \sigma_A^2/R$ puede aproximarse calculando $\hat{\sigma}_A^2 = (R - 1)^{-1} \sum_{i=1}^{i=R} (a_i - a_{(\cdot)})^2$, la varianza “muestral”¹. De las anteriores fórmulas surge que el número de repeticiones necesario para asegurar una precisión de ε con una confianza de $1 - \alpha$ es:

$$R > (Z_{1-\alpha/2})^2 \times \hat{\sigma}_A^2 / \varepsilon^2. \quad (1)$$

Con el “tamaño de muestra” determinado en 1 podemos asegurar la precisión y confianza requeridas para asegurar $P(|a_{(\cdot)} - \mu_A| > \varepsilon) \leq \alpha$. Podemos hacer lo mismo para determinar un valor de R que asegure la precisión y confianza requeridas para asegurar $P(|b_{(\cdot)} - \mu_B| > \varepsilon) \leq \alpha$ y luego tomar el supremo de ambos números.

Ejemplo.

Considere el siguiente código escrito en Matlab:

¹Para tener una primer aproximación a $\hat{\sigma}_A^2$ puede hacerse un número inicial de repeticiones (100, 500, ...).

```

r=1000
a=zeros(r,1);
b=zeros(r,1);
for i=1:r
    c(i)=unidrnd(6);
    v(i)=unidrnd(6);
    a(i)=(c(i)<=5);
    b(i)=(c(i)<=v(i));
end
abar=mean(a)
asig=var(a)
bbar=mean(b)
bsig=var(b)

```

El siguiente cuadro presenta los valores exactos y los valores estimados de los parámetros usando un número inicial de 1000 repeticiones.

Parámetro	Valor exacto	Valor estimado
μ_A	0,8333	0,8250
σ_A^2	0,1389	0,1445
μ_B	0,5833	0,5770
σ_B^2	0,2431	0,2443

Cuadro 1: *Valores aproximados y simulados*

Los valores de R necesarios para asegurar una precisión de 0.001 con una confianza del 99 % para μ_A y μ_B son 921.591 y 1:612.951 respectivamente, por lo que en este caso deberíamos hacer al menos 1:612.951 iteraciones.

Otros parámetros.

En general la información producida por simulaciones Monte Carlo son promedios o funciones de promedios calculadas con R repeticiones del algoritmo. El siguiente cuadro presenta las fórmulas de las varianzas aproximadas de los estimadores por tipo de estimador. Todas las sumas que aparecen son para $i = 1, \dots, R$.

En el caso que los parámetros a presentar en los cuadros sean cuantiles u otras funciones que no puedan ser expresadas como funciones de promedios, las fórmulas de las varianzas del cuadro anterior no son aplicables.

Item reportado	Estimador	Varianza estimada
Media	$a_{(\cdot)} = R^{-1} \sum a_i$	$R^{-2} \sum (a_i - a_{(\cdot)})^2$
Raíz cuadrada de media	$(a_{(\cdot)})^{1/2}$	$4^{-1} (a_{(\cdot)})^{-1} R^{-2} \sum (a_i - a_{(\cdot)})^2$
Ratio de medias	$r = a_{(\cdot)} (b_{(\cdot)})^{-1}$	$(b_{(\cdot)})^{-2} R^{-2} \sum (a_i - r b_i)^2$
Raíz cuadrada de ratio	$r^{1/2}$	$4^{-1} r^{-1} (b_{(\cdot)})^{-2} R^{-2} \sum (a_i - r b_i)^2$

Cuadro 2: *Valores aproximados* de las varianzas de los parámetros simulados

Peña Sánchez de Rivera, Daniel (2001). “Deducción de distribuciones: el método de Montecarlo”, en Fundamentos de Estadística. Madrid: Alianza Editorial. ISBN 84-206-8696-4.