Prueba de aleatoriedad usando integración por el método Monte Carlo

Dr. A. Giovanni Ramírez García*

Guatemala, 26 de abril de 2021

1. Integración por el método del punto medio

El objetivo de este método es encontrar el área bajo la curva $f(x)$ y limitada por las rectas $x = a$ y $x = b$ dada por la integral

$$\int_a^b dx f(x). \tag{1}$$

El teorema del valor medio establece que si $\bar{f}$ es el valor medio de f en $[a, b]$, entonces el área dada por (1) es

$$\int_a^b dx f(x) = (b-a)\bar{f}, \tag{2}$$

esta relación se suele usar para calcular el valor medio

$$\bar{f} = \frac{1}{b-a}\int_a^b dx f(x).$$

La definición del valor medio dada por (2) permite el uso de un método numérico de integración para aproximar $\bar{f}$ tomando un promedio de f en un conjunto de n puntos $\{x_0, \dots, x_{n-1}\}$

$$\bar{f} \approx \frac{1}{n}\sum_{i=0}^{n-1} f(x_i). \tag{3}$$

Se puede aprovechar la libertad en cómo elegir el conjunto de puntos $\{x_i\}$. Por ejemplo, si se toman uniformemente espaciados en $[a, b]$ se tendría

$$x_i = a + ih + \frac{1}{2}h, \quad \forall i = 0, \dots, n-1,$$

donde $h = (b-a)(n-1)^{-1}$, que corresponde a la regla del punto medio para integración numérica[1, 2]. Es intuitivo que mientras más puntos se tomen mejor será la aproximación de $\bar{f}$. Haciendo un análisis más detallado[3] se puede ver que el error de aproximación decrece con n^{-2}.

También se puede elegir el conjunto $\{x_i\}$ de una forma distinta, por ejemplo dada por

$$x_i = a + ih, \quad \forall i = 0, \dots, n-1,$$

donde $h = (b-1)(n-2)^{-1}$. Esta distribución tiene la ventaja de que los puntos inicial y final coinciden con los extremos del intervalo. Sin embargo, el error de aproximación en este caso decrece sólo con n^{-1}.

2. Integración usando puntos aleatorios

Otra opción para elegir el conjunto de puntos $\{x_i\}$ es usar puntos aleatorios uniformemente distribuidos en $[a, b]$. A este método se le llama integración por el método Monte Carlo[3]. El análisis de error muestra que el error decrece con $n^{-1/2}$, es decir es menos preciso que el método de punto fijo. Sin embargo, cuando se aplica este método a funciones definidas sobre el plano $\mathbb{R}^2$ o sobre

*Escuela de Ciencias Físicas y Matemáticas, USAC.

el espacio $\mathbb{R}^3$, el la integración numérica con el método Monte Carlo es más efectiva[3].

Desde el punto de vista de la teoría de probabilidad, las integrales como la de la ecuación (1) representan el valor esperado de una variable aleatoria, o también llamada la esperanza matemática de $f(x)$ si x es una variable aleatoria uniformemente distribuida en $[a, b]$.

La integral con el método de Monte Carlo se estima como el promedio de las muestras aleatorias. Si se considera que la función de densidad de probabilidad de una variable aleatoria $X \in [a, b]$ está dada por

$$p(x) = \begin{cases} (b-a)^{-1}, & x \in [a, b], \\ 0, & \text{en otra parte.} \end{cases}$$

Entonces, la esperanza matemática está dada por

$$\begin{aligned} E(f(X)) &= \int_{-\infty}^{\infty} dx f(x) p(x) \\ &= (b-a) \int_a^b dx f(x), \end{aligned}$$

ahora bien, el valor esperado para una muestra grande de números uniformemente distribuidos se aproxima a

$$E(f(X)) \approx \frac{1}{n} \sum_{i=0}^{n-1} f(x_i),$$

entonces, finalmente se tiene el método de integración por el método Monte Carlo

$$\int_a^b dx f(x) \approx \frac{b-a}{n} \sum_{i=0}^{n-1} f(x_i). \qquad (4)$$

3. Prueba de aleatoriedad

Una aplicación del método de integración por Monte Carlo es la prueba de aleatoriedad de una variable aleatoria con una distribución de densidad de probabilidad uniforme. Para poder hacer esta prueba a otras distribuciones de densidad de probabilidad hay que extender el método de integración por Monte Carlo para hacer integrales usando el teorema del valor medio ponderado.

Consideremos nuevamente la ecuación (4). Si ahora elegimos una función fácil de integrar, por ejemplo $f(x) = x$, de modo que al hacer la integral

$$\int_a^b dx(x) = \left.\frac{x^2}{2}\right|_a^b = \frac{(b-a)^2}{2}.$$

y tomar la aproximación por el método de integración por Monte Carlo se tiene

$$\frac{1}{n} \sum_{i=0}^{n-1} x_i \approx \frac{b-a}{2},$$

que se cumple si y sólo si la variable aleatoria x_i está uniformemente distribuida en $[a, b]$.

Funciones más complejas necesitan más puntos para converger al valor exacto de la integral. Por ejemplo, la aproximación de $f(x) = x$ con un conjunto de 10^6 números tiene un error relativo de 10^{-3}, mientras que $f(x) = \sin(x)$ requiere muchos más puntos.

Referencias

[1] R. Burden, J. Faires, and A. Burden, *Numerical Analysis*. Cengage Learning, 2015.

[2] H. Cohen, *Numerical Approximation Methods: $\Pi \approx$ 355/113*. Springer, 2011.

[3] H. Langtangen, *A Primer on Scientific Programming with Python*. Texts in Computational Science and Engineering, Springer Berlin Heidelberg, 2014.