

Integración por el método de Monte Carlo

Los métodos de Monte Carlo fueron originados en Los Alamos National Laboratory poco después de la Segunda Guerra Mundial. Si bien hubo algunos antecedentes como los de Comte de Buffon que arrojaba una aguja sobre un tablero con una cuadrícula o los de Kelvin que estudiaba la cinética de los gases, estos métodos aleatorios se desarrollaron e impusieron con la aparición de la computadora. Una vez concluida la primera computadora electrónica de Estados Unidos (ENIAC), los científicos de Los Alamos consideraron su uso para el diseño de armas termonucleares. En 1946 Stanislaw Ulam sugirió el uso de un muestreo aleatorio para simular el recorrido de neutrones y John von Neumann desarrolló una propuesta detallada a principios de 1947. Estas simulaciones en pequeña escala fueron decisivas para completar el proyecto. Metropolis and Ulam (1949) publicaron un paper en el que describieron sus ideas, que dispararon gran cantidad de investigaciones en la década del 50. El nombre de Monte Carlo proviene de la ciudad homónima en Mónaco, famosa por sus casinos.

Una aplicación importante de este método es al cálculo de integrales, especialmente de alta dimensión, ya sea porque no es posible calcularlas en forma exacta o es muy difícil su cómputo con el grado de precisión deseado. La idea de la *integración por Monte Carlo* es evaluar la integral

$$I = \int_{\Omega} g(x) dx$$

usando muestreo aleatorio. En su forma más elemental esto se realiza muestreando en forma independiente n puntos X_1, \dots, X_n cada uno con densidad conveniente f con dominio en Ω y luego calculando

$$I_n^1 = \frac{1}{n} \sum_{i=1}^n \frac{g(X_i)}{f(X_i)}$$

Por ejemplo, si el dominio de integración es el hipercubo unidad de dimensión q , es decir $\Omega = [0, 1]^q$, podemos elegir X_i independientes y con distribución uniforme. En este caso resulta

$$I_n^2 = \frac{1}{n} \sum_{i=1}^n g(X_i)$$

Una posibilidad podría ser buscar una densidad f de manera tal que $\frac{g(x)}{f(x)}$ fluctúe menos que g y de esta manera reducir la varianza de I_n^2 . Esto se conoce como *importance sampling*.

El método de Monte Carlo tiene las siguientes ventajas:

- i) es muy simple
- ii) converge a velocidad $\frac{1}{\sqrt{n}}$ en cualquier dimensión independientemente de la suavidad de la función a integrar. Esto no es así para otros métodos muy populares
- iii) es versátil y por lo tanto puede usarse aún para dominios muy generales

iv) mediante *importance sampling* pueden integrarse funciones con singularidades.

Preguntas

- a) ¿Cuál es la esperanza de I_n^1 ?
- b) Encontrar una expresión para la varianza de I_n^1 y observar que si $E \left[\left(\frac{g(x)}{f(x)} \right)^2 \right] < \infty$ el desvío standard de I_n^1 tiende a 0 con velocidad $\frac{1}{\sqrt{n}}$.
- c) Supongamos que existe una constante M tal que $\left| \frac{g(x)}{f(x)} \right| < M$. ¿Cuán grande debería ser n si quisiéramos que la diferencia entre I_n^1 e I fuera a los sumo 0.01 con probabilidad mayor a 0.90?
- d) ¿Es I_n^1 un estimador consistente de I ? Justificar cuidadosamente.

Integración por el método de rechazo

Existen diversas técnicas de Monte Carlo para la evaluación de integrales, entre las más sencillas encontramos la técnica de rechazo, que se basa en la siguiente idea muy intuitiva, que presentaremos con un ejemplo.

Supongamos que queremos medir el área de un lago L con lo único que tenemos a mano. Primero podríamos recorrer la zona determinando una región rectangular R que contenga al lago como en la Figura 1 y cuyas medidas podamos definir fácilmente (por ejemplo con la longitud de nuestros pasos, o una varilla) como en el gráfico. De esta forma podemos calcular el área de R , $\mathcal{A}(R)$. Ahora escogemos un número N de piedritas, las arrojamamos aleatoriamente sobre R y contamos el número de piedritas que se sumergieron en el lago, digamos N_L . Como el número de piedritas que se hundieron es proporcional al área de L , $\mathcal{A}(L)$, tenemos que:

$$\begin{aligned} \frac{N_L}{N} &= \frac{\mathcal{A}(L)}{\mathcal{A}(R)} \\ \mathcal{A}(L) &= \frac{N_L}{N} \mathcal{A}(R) \end{aligned}$$

Con esta idea, si queremos calcular la integral de la función g en una región R como en la Figura 2, elegimos puntos aleatoriamente en la región R y la integral de la función g la estimamos como el área de R multiplicada por la fracción de puntos que caen por debajo de la curva g .

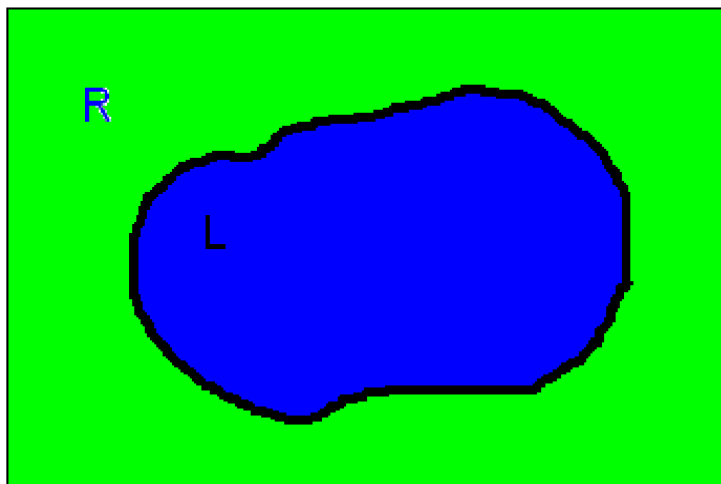


Figura 1: Superficie del lago

Preguntas

Como sabemos el área del círculo unitario es π , es decir

$$\int \int_{\text{circulo}} dx dy = \pi$$

- e) Encuadrando el círculo unitario en el cuadrado de vértices $(1, 1), (-1, 1), (1, -1)$ y $(-1, -1)$, estimar π mediante el método del rechazo usando $n = 100, 1000$ y 10000 . Comparar los resultados.
- f) ¿Qué n sería necesario para obtener una precisión de 2 dígitos con probabilidad mayor a 0.90? Calcular el estimador para una muestra del tamaño mínimo hallado.
- g) Replicar 5000 veces la estimación basada en muestras de tamaño $n = 10$, obteniendo 5000 estimaciones de π . Utilizando las técnicas descriptivas que conoce: boxplot, histograma y qq-plot, estudiar la distribución del estimador.
- h) Repetir g) con $n = 20, 50$ y 100 y comparar las distribuciones obtenidas. ¿Es esperable el comportamiento que observa a medida que n crece? ¿Por qué?

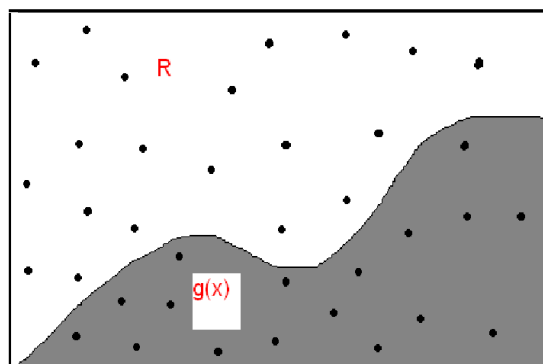


Figura 2: $\int g(x) dx$