Enunciados prácticas y soluciones selectas

2. Simulación de Monte Carlo - Decisiones

2.4. Integración de Monte Carlo

Simulación de Sistemas

3007331

Práctica 4 - Método aceptación rechazo e integración Monte Carlo

Esta práctica tiene dos partes. En la primera se implementa el método de aceptación/ rechazo para generar una muestra de observaciones de una variable aleatoria. En la segunda se implementa un método de Monte Carlo basado en una idea similar a la de aceptación rechazo para calcular el valor de una integral definida.

Parte 1 - Aceptación / Rechazo

Objetivos de aprendizaje

- Crear un programa en python o R para generar variables aleatorias que sigan una función de densidad de probabilidad dada usando el método de aceptación-rechazo.
- Generar y analizar histogramas de las variables aleatorias generadas y verificar que siguen la distribución deseada.

Actividad parte 1:

- Escribir un programa en Python para generar observaciones de una variable aleatoria x con la función de densidad de probabilidad $f(x) = 20x(1-x)^3, \quad 0 < x < 1$
- Generar N observaciones y graficar el histograma de la distribución de la muestra. Comparar con la gráfica de la fdp.
- Calcular cuántas observaciones se rechazan para generar las N observaciones anteriores.

Parte 2. Integración de Monte Carlo

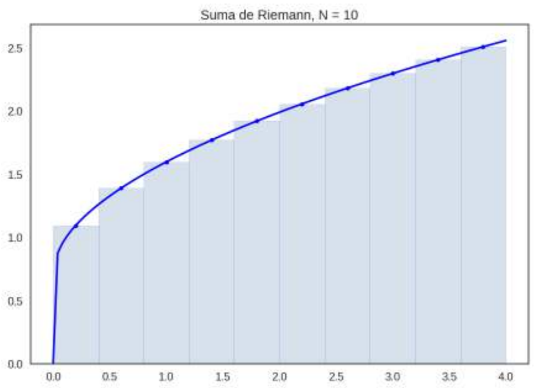
Objetivos de aprendizaje

- Construir una simulación de Monte Carlo para aproximar la solución a un problema que pueda formularse como una integral definida.
- Evaluar la exactitud y precisión de la simulación por medio del cálculo de errores estándares e intervalos de confianza.

Integración Numérica

La integral definida $I = \int_a^b f(x)\,dx$ de una función $f(x)$ sobre un intervalo finito $[a, b]$ es un número que representa el área bajo la curva de $f(x)$ entre a y b . Si se conoce una antiderivada o primitiva $F(x)$ de $f(x)$ tal que $F'(x) = f(x)$, el valor de la integral se calcula como $I = \int_a^b f(x)\,dx = F(b) - F(a)$.

No es fácil encontrar la antiderivada de muchas funciones y en tales casos, es necesario aproximar el valor de la integral. Los métodos numéricos más simples dividen el intervalo $[a, b]$ en N subintervalos pequeños y aproximan el área bajo la curva en cada pequeño intervalo como el área de un trapecio o un rectángulo, dependiendo del método elegido. Al sumar las áreas de todos los rectángulos (o trapecios), se tiene un valor aproximado de la integral. En la figura se ilustra el método sumando las áreas de rectángulos.



Para intervalos lo suficientemente pequeños, los métodos como el trapezoidal o la regla de Simpson pueden aproximar el valor de una integral definida con precisión aceptable. Sin embargo, las técnicas numéricas analíticas son poco eficientes para calcular las integrales de funciones que no tienen derivadas continuas y especialmente las de funciones en más de una dimensión.

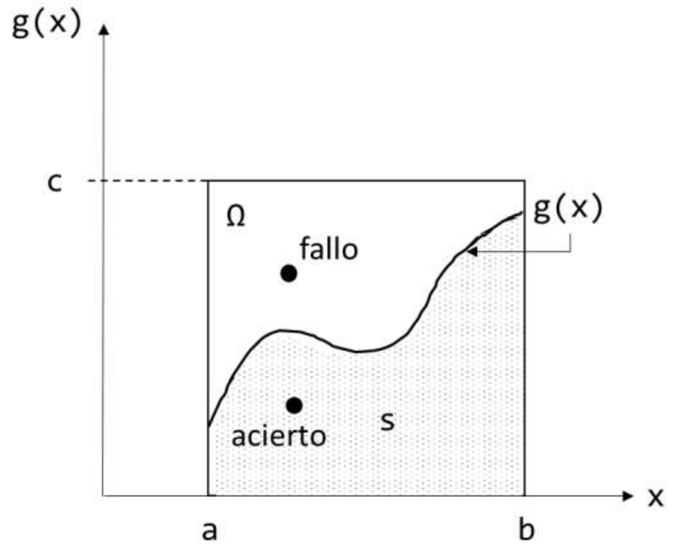
Con frecuencia, en estos casos es más conveniente usar un método de Monte Carlo que, aunque menos preciso que las fórmulas de cuadratura convencionales, es más simple de usar y más eficiente computacionalmente.

Integración de Monte Carlo

Cada integral se puede representar como un valor esperado (un parámetro) y estimar una integral por el método de Monte Carlo equivale a estimar un parámetro desconocido. El método visto en esta práctica, acertar/errar, se basa en la interpretación geométrica de una integral como un área. A continuación se detalla el método siguiendo la explicación en (Rubinstein,1981).

Método Acertar/errar

Se va a calcular una integral unidimensional. Por simplicidad, se asume que el integrando $g(x)$ es acotado entre a y b:
$0 \leq g(x) \leq c, \quad a \leq x \leq b$



Sea Ω el rectángulo definido por a,b,c: $\Omega = (x, y) : a \leq x \leq b, 0 \leq y \leq c$.

La probabilidad de que un vector (X, Y) distribuido uniformemente sobre el rectángulo Ω caiga bajo la curva $g(X)$ es

$p = \frac{areaS}{area\Omega} = \frac{\int_a^b g(x)dx}{c(b-a)} = \frac{I}{c(b-a)}$

Supongamos que se pueden generar N vectores aleatorios $(X_1, Y_1), (X_2, Y_2), \ldots (X_N, Y_N))$. El parámetro p se puede estimar como

$\hat{p} = \frac{N_H}{N}$ donde N_H es el número de ocasiones en que $g(X_i) \geq Y_i, \quad i = 1, 2, \ldots, N$ es decir, el número de aciertos y $N - N_H$ el número de fallos.

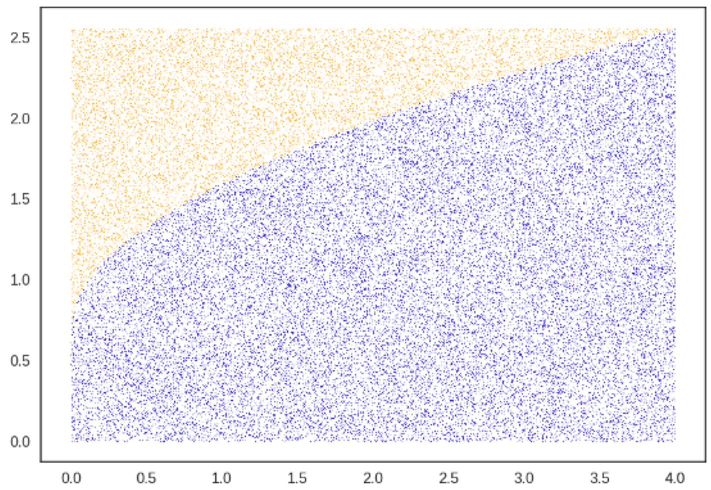
De lo anterior, la integral I se puede estimar como $I \approx \theta_1 = c(b-a)\frac{N_H}{N}$. Cada vector generado tiene una probabilidad p de ser un acierto y es un ensayo de Bernoulli.

- θ_1 es un estimador no sesgado de I porque $E(\theta_1) = c(b-a)E(\frac{N_H}{N}) = c(b-a)\frac{E(N_H)}{N} = c(b-a)p = I$
- La varianza de θ_1 es $var\theta_1 = \frac{I}{N}[c(b-a) - I]$
- Para estimar θ_1 con un error ε y un nivel de confianza α se requieren $N \geq \frac{(1-p)p[c(b-a)]^2}{(1-\alpha)\varepsilon^2}$ ensayos.
- Para N lo suficientemente grande, el intervalo de confianza con nivel $1 - 2\alpha$ para I está dado por $\theta_1 \pm z_\alpha \frac{[\hat{p}(1-\hat{p})]^{1/2}(b-a)c}{N^{1/2}}$

Algoritmo

1. Generar una secuencia de N pares de números $(U_1, U_1'), (U_2, U_2'), \ldots (U_N, U_N')$ donde cada U_i, U_i' sigue una distribución uniforme entre 0 y 1 $U_i, U_i' \sim U[0, 1)$.
2. Calcular $X_i = a + U_i(b-a)$ y $g(X_i)$ $i = 1, 2, \ldots, N$.
3. Contar el número de casos N_H en que se cumple que $g(X_i) > cU_i'$.
4. Estimar la integral I como $\theta_I = c(b-a)\frac{N_H}{N}$.

En la Figura 2 se ilustra el resultado de los aciertos (azul) y los fallos (naranja) para la misma función de la Figura 1



Actividad de la segunda parte

1. Estimar el valor de $I = \int_0^1 sen^2(\pi x)dx$ usando el método de acierto/falla y calcular un intervalo de confianza para esta estimación.
2. Comparar el valor obtenido con la solución que arroja una aplicación especializada, como Mathematica.

3. Graficar los resultados: valores aceptados y función a integrar.

ENTREGABLES

Parte 1:

1. Código en python.
2. Histograma de las observaciones de la muestra, calculado usando el método descrito en el libro de ejercicios resueltos. Es decir, calculando el número de clases, y contando las observaciones por clase.
3. Respuesta a las preguntas

Parte 2:

1. Diagrama de flujo del procedimiento para calcular la integral
2. Respuesta a las preguntas