Métodos de Monte Carlo

Entrega 7 - Grupo 01

(Unidad 5, Sesión 14) Ejercicio 14.1

Integrantes

Nombre	CI	Correo
Nadia Martínez	5.141045-0	nadia.martinez.petrone@fing.edu.uy
Germán Ouviña	4.823.566-1	german.ouvina@fing.edu.uy

Índice

1. Problema . . . 3

2. Solución . . . 4

2.1. Descripción . . . 4

2.2. Código . . . 4

2.3. Pseudocódigo . . . 5

3. Experimentación . . . 6

3.1. Especificaciones . . . 6

3.2. Resultados . . . 6

1. Problema

Ejercicio 14.1 (GRUPAL):

Partiendo de uno de los códigos elaborados para resolver el ejercicio 6.2, utilizar el método de muestreo estratificado para calcular la integral de la función $x_1x_2^2x_3^3x_4^4x_5^5$ sobre el hipercubo J^m de dimensión $m = 5$ en base a 10^6 iteraciones. Calcular media, desviación estándar y un intervalo de confianza de nivel 95 %. Comparar con los resultados obtenidos con el código del ejercicio 6.2.

Sugerencia: definir 5 estratos, en función del valor de x^5, tomando los siguientes intervalos:

$$[0; 0{,}72), \quad [0{,}72; 0{,}83), \quad [0{,}83; 0{,}90), \quad [0{,}90; 0{,}95), \quad [0{,}95; 1].$$

Hacer dos experimentos, uno tomando $10^6/5$ iteraciones en cada estrato, otro tomando una cantidad de iteraciones proporcional a la probabilidad de cada estrato.

2. Solución

2.1. Descripción

Para este ejercicio, se utilizó como base el código realizado para resolver el ejercicio 6.2, el cual fue parte de la tercera entrega del curso. Este código se modificó para una vez más estimar la integral de una función sobre un hipercubo, pero ahora utilizando el método de Monte Carlo con muestreo estratificado, con el objetivo de reducir la varianza y mejorar la eficiencia del algoritmo en general.

Para implementar este método, en lugar de muestrear puntos del dominio uniformemente, se divide el dominio de la variable x_5 en cinco estratos disjuntos, como se indica en la sugerencia de la letra. Luego se realizan dos experimentos, uno con asignación uniforme de muestras entre los estratos, y otra con asignación proporcional al tamaño de cada estrato. En ambos casos, las coordenadas x_1, x_2, x_3 y x_4 se muestrean uniformemente en el intervalo $[0, 1)$ como antes.

2.2. Código

El código para solucionar el problema se encuentra adjunto en un archivo *.zip* junto al informe, y se encuentra disponible en el repositorio `https://github.com/gouvina-fing/fing-mmc2025`.

El mismo fue desarrollado en *Python*, cuenta con instrucciones de como instalarlo y correrlo, y ofrece la posibilidad de generar resultados para ambos tipos de estratificación a la vez.

En la siguiente sección se proporciona un bosquejo del código en formato de pseudocódigo, con una explicación escueta de los cambios introducidos.

2.3. Pseudocódigo

Procedimiento EstimacionMonteCarloIntegral (int n, funcion f, real δ, int m)
Parámetros de entrada:

- n - tamaño de la muestra
- f - función a calcular su integral
- δ - nivel de confianza
- m - cantidad de estratos

Parámetros de salida:

- $\hat{X}$ - estimador de la integral
- $\hat{V}_f$ - estimador de la varianza de la función f
- ω_1, ω_2 - intervalo de confianza para $(1-\delta)$

Pseudocódigo:

1. $X, \hat{X}, \hat{V}_f = 0$
2. Calculo de estratos según método de estratificación:
   a) Para el método uniforme, 5 estratos de tamaño $10^6/5$
   b) Para el método proporcional, 5 estratos de tamaño $n*(b-a)$ siendo a, b el comienzo y fin de cada estrato
3. For $i = 1, ..., m$ do
   a) $a_i, b_i =$ comienzo del estrato i
   b) $n_i =$ tamaño del estrato previamente calculado i
   c) $p_i = b_i - a_i$
   d) For $j = 1, ..., n_i + 1$
      1) Sortear coordenadas de un punto p siguiendo la distribución $U(0,1)$ por coordenada para x_1, x_2, x_3 y x_4
      2) Sortear coordenada $x_5 = U(a_i, b_i)$
      3) Evaluar $f(p)$
      4) Acumular $X_i = X_i + f(p)$
      5) Si $i > 1 \rightarrow \hat{V}_{f_i} = \hat{V}_{f_i} + (1 - \frac{1}{i})(f(p) - \frac{X}{i-1})^2$
   e) $X_i = X * p_i / n_i$
   f) $V_{f_i} = V_{f_i} * p_i^2/(n_i - 1)$
4. $\hat{X} = \sum_1^m X_i$
5. $\hat{V}_f = \sum_1^m \hat{V}_{f_i}$
6. $\omega_1, \omega_2 = I_1(\hat{X}, \sqrt{\hat{V}_X}, \delta), I_2(\hat{X}, \hat{V}_X, \delta)$

El anterior pseudocódigo bosqueja un método de monte carlo para el cálculo de la integral de una función, siguiendo las pautas del pseudocódigo presentado en clase. Con respecto al código entregado anteriormente, el cambio se encuentra en la generación de $m = 5$ estratos acorde a los intervalos planteados en la consigna. El uso de estratos requiere calcular estimadores y varianzas para cada uno de ellos y luego acumularlos en uno, lo cual impacta en el uso de dos bucles anidados (el superior para los estratos, el inferior para sus tamaños de muestras). El resto del algoritmo no cambia.

3. Experimentación

3.1. Especificaciones

A continuación, se desglosan las especificaciones técnicas utilizadas durante la ejecución:

- **Procesador:** 11th Gen Intel(R) Core(TM) i5-11400F @ 2.60GHz
- **Memoria RAM:** 16.0 GB
- **Sistema Operativo:** Windows 10
- **Semilla:** 42
- **Tamaños de muestra:** 10^6

3.2. Resultados

La consigna de este ejercicio cuenta con dos partes en lo que refiere a la experimentación: en primera instancia, ejecutar el algoritmo tomando un tamaño de muestra uniforme para cada estrato (en este caso, $10^6/5$ debido a que el tamaño de muestra base es 10^6 y se divide en 5 estratos). En segunda instancia, se ejecuta el algoritmo tomando una cantidad de iteraciones proporcional a la probabilidad de cada estrato, la cual viene modelada por la consigna. En ambos casos se calcula la media, desviación estándar e intervalo de confianza al 95 % utilizando la aproximación normal.

A continuación, se adjunta la tabla comparativa con los resultados obtenidos para ambos tipos de muestreo, comparando a su vez con los resultados obtenidos en la entrega original sin estratificar la muestra:

Método de muestreo	**Estimador**	**Varianza**	**Desviación**	**IdeC (Normal)**	**Tiempo (s)**
Sin estratificar	0.0013932313	$9{,}3605 \times 10^{-5}$	$9{,}6750 \times 10^{-6}$	0.0013742687 0.0014121939	10.51
Estratificación uniforme	0.0013881813	$3{,}5540 \times 10^{-11}$	$5{,}9615 \times 10^{-6}$	0.0013764968 0.0013998657	9.80
Estratificación proporcional	0.0013836696	$8{,}9653 \times 10^{-11}$	$9{,}4685 \times 10^{-6}$	0.0013764968 0.0013998657	9.95

A partir de estos resultados, se observa que el muestreo estratificado ofrece una mejora significativa respecto al método de Monte Carlo sin estratificar utilizado en el ejercicio 6.2, al menos en lo que refiere a la varianza. Ambas variantes del muestreo estratificado logran estimadores similares al método clásico, pero con una varianza notablemente menor, lo cual potencialmente afecta la eficiencia del algoritmo. En este caso particular, la varianza se reduce en varios órdenes de magnitud, pasando de un valor del orden de 10^-5 en el método no estratificado a valores del orden de 10^-11 con el muestreo estratificado; mejorando la precisión sin necesidad de incrementar la cantidad de muestras ni el tiempo de ejecución (el cual se mantuvo en ordenes similares, disminuyendo menos de 1 segundo en comparación). Además, el intervalo de confianza es más estrecho, lo cual indica una mayor confiabilidad en la estimación.

La ventaja de utilizar el muestreo estratificado es que controla mejor la variabilidad de la función en las distintas regiones del dominio. En el caso del método utilizado en el ejercicio 6.2, las muestras se generan de forma aleatoria en todo el espacio, provocando que puedan tomarse más muestras en zonas poco representativas. Si la función tiene mas variación en ciertas partes del dominio, esto va a aumentar la varianza del estimador. El muestreo estratificado divide el dominio en regiones ”homogéneas” (denominadas estratos) y toma muestras dentro de cada una. De esta forma se garantiza que todas las regiones relevantes estén representadas proporcionalmente en la muestra final, siempre y cuando, los estratos se elijan con criterio. Como se estima la integral por separado en cada estrato y luego se combinan los resultados, la variabilidad dentro de cada grupo es menor, obteniendo así una varianza total más baja.

En este caso, como la función es más "sensible" en x_5 (porque está elevada al grado 5), las regiones donde esta variable es grande aportan valores mucho mayores. Al estratificar específicamente sobre x_5, se asegura el cubrimiento de forma homogénea tanto en regiones donde la función es más estable como en aquellas donde crece más rápido.