

MÉTODOS DE MONTE CARLO

ENTREGA 2

(Unidad 2, Sesiones 3 y 4) Ejercicios 3.1 y 4.1

Integrantes

Nombre	CI	Correo
Germán Ouviaña	4.823.566-1	german.ouvina@fing.edu.uy

Índice

1. Problema	3
2. Solución	4
2.1. Descripción	4
2.1.1. Ejercicio 3	4
2.1.2. Ejercicio 4	4
2.2. Pseudocódigo	5
2.3. Código	5
3. Experimentación	6
3.1. Especificaciones	6
3.2. Resultados	6
3.2.1. Ejercicio 3	6
3.2.2. Ejercicio 4	8

1. Problema

Se desea estimar el volumen de una región R de $[0, 1]^6$ definida por todos los puntos de la hiper-esfera de centro $(0,45, 0,5, 0,6, 0,6, 0,5, 0,45)$ y radio $0,35$, que además cumplan las restricciones siguientes: $3x_1 + 7x_4 \leq 5$; $x_3 + x_4 \leq 1$; $x_1 - x_2 - x_5 + x_6 \geq 0$

Entrega 2: Ejercicio 3.1 (INDIVIDUAL):

1. Implementar un programa que reciba como parámetro la cantidad de replicaciones n a realizar, y emplee Monte Carlo para calcular (e imprimir) la estimación del volumen de R , y la desviación estándar de este estimador. Incluir código para calcular el tiempo de cálculo empleado por el programa. Utilizar el programa con $n = 10^4$ y luego con $n = 10^6$ para estimar el volumen de R . Discutir si los dos valores obtenidos parecen consistentes.
2. Como forma de validar el programa, eliminar las restricciones adicionales de desigualdad, y comparar el volumen calculado por Monte Carlo con $n = 10^6$ con el valor exacto del volumen de una hiperesfera de dimensión 6, $\pi^3 r^6 / 6$. Discutir también la relación de este valor con el obtenido en la parte a.

Entrega 2: Ejercicio 4.1 (INDIVIDUAL):

1. Comparar y discutir la dependencia de los criterios de peor caso n_C, n_N, n_H frente a los parámetros ϵ y δ
2. Calcular n_C, n_N, n_H para $\epsilon = 0,01, \delta = 0,001, 0,01, 0,05$

2. Solución

2.1. Descripción

2.1.1. Ejercicio 3

Este ejercicio pide calcular el volumen de una región compuesta por la intersección de una hiperesfera y otras restricciones en relación a las coordenadas de los puntos. Al igual que en la entrega anterior, el programa implementado parte de la base de un método de monte carlo básico, especificando pasos extra para comprobar si los puntos generados se hallan dentro de la región o no. En las secciones 2.2 y 2.3 se dan más detalles sobre el programa, y en la sección 3.2.1 se discuten los resultados obtenidos.

2.1.2. Ejercicio 4

Este ejercicio pide un estudio básico sobre los criterios de peor caso para el cálculo de intervalos de confianza en métodos de monte carlo.

Dado un error ϵ y un intervalo de confianza $1 - \delta$, existen tres derivaciones que permiten inferir el tamaño de la muestra en el peor caso para mantener el error dentro de los parámetros del intervalo. Dichas derivaciones parten de los siguientes métodos:

- **Desigualdad de Chebyshev**
 $n_C(\epsilon, \delta) = \lceil 1/(4\delta\epsilon^2) \rceil$
- **Teorema Central del Límite**
 $n_N(\epsilon, \delta) = \lceil (\Phi^{-1}(1 - \delta/2)/2\epsilon)^2 \rceil$ siendo Φ^{-1} la función inversa a la normal $N(0, 1)$
- **Teorema de Hoeffding**
 $n_H(\epsilon, \delta) = \lceil 2\ln(2/\delta)/(4\epsilon^2) \rceil$

En la sección 3.2.2 se discuten los resultados obtenidos para los calculos en base a los distintos valores de ϵ y δ pedidos.

2.2. Pseudocódigo

Procedimiento EstimacionMonteCarloVolumen (int n , region R , real \hat{X} , real \hat{V})

Parámetros de entrada: n (tamaño de la muestra), R (región a calcular el volumen)

Parámetros de salida: \hat{X} , \hat{V} (estimadores de volumen y varianza)

1. $\hat{X} = 0$
2. $\hat{V} = 0$
3. For $i = 1, \dots, n$ do
 - a) Sortear coordenadas de un punto X_j siguiendo la distribución $U(0, 1)$ por coordenada
 - b) Si $X_j \in R \rightarrow \hat{X} = \hat{X} + 1$
4. $\hat{X} = \hat{X}/n$
5. $\hat{V} = \hat{X}(1 - \hat{X})/(n - 1)$

El anterior pseudocódigo bosqueja un método de monte carlo sencillo para el cálculo de volumen de una región, siguiendo las pautas del pseudocódigo presentado en clase. Concretamente, el paso 3b es donde se halla la diferencia en esta implementación, ya que el chequeo específico utiliza los parámetros dados en la letra para la hiperesfera y las distintas restricciones que definen la región.

2.3. 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 tablas y gráficas comparativas siguiendo la consigna del ejercicio 3, automáticamente ejecutando el método para los distintos valores de n mencionados y cortando la ejecución cuando el tiempo supera los 60 segundos o las 10^7 iteraciones para evitar overflows. También existe un programa aparte para el cálculo de los tamaños de muestra pedidos en el ejercicio 4, el cual genera una tabla comparativa para los distintos valores de δ pedidos.

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, 10², 10³, 10⁴, 10⁵, 10⁶, 10⁷

3.2. Resultados

3.2.1. Ejercicio 3

A continuación, se muestra una tabla comparativa de los resultados obtenidos al ejecutar el método **con restricciones** para distintos tamaños de muestra. Se destacan **en rojo** los pedidos en la consigna:

N° de iteraciones	\hat{X}	\hat{V}	Tiempo (s)
10	0	0	0
10 ²	0	0	0.000998
10 ³	0	0	0.012001
10 ⁴	0.000400	0.000199969	0.116003
10 ⁵	0.000250	0.000050983	1.214703
10 ⁶	0.000281	0.000001690	11.955126

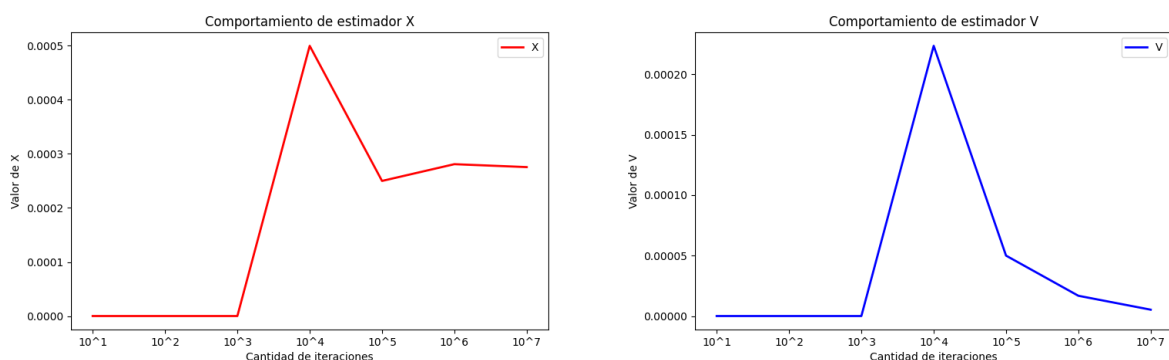


Figura 3.1: Comportamiento de estimadores en cálculo de volumen CON restricciones

Tanto de la tabla como de las gráficas se puede observar el comportamiento esperado: el estimador de volumen oscila hasta mantenerse alrededor de un valor, mientras que el estimador de la varianza disminuye con el aumento del tamaño de muestra. Es interesante observar que recién al emplear un tamaño de muestra de 10⁴ se observa un valor en el estimador de volumen: con menos iteraciones, los puntos sorteados parecen no ser suficientes para entrar en la región y el estimador queda en cero.

A continuación, se muestra una tabla comparativa de los resultados obtenidos al ejecutar el método **sin restricciones** para distintos tamaños de muestra. Se destacan **en rojo** los pedidos en la consigna:

N° de iteraciones	\hat{X}	\hat{V}	Tiempo (s)
10	0	0	0
10^2	0	0	0.001001
10^3	0.006000	0.002443	0.012755
10^4	0.008400	0.0009127032	0.126652
10^5	0.009310	0.0003037007	1.269264
10^6	0.009440	0.000097	12.657965

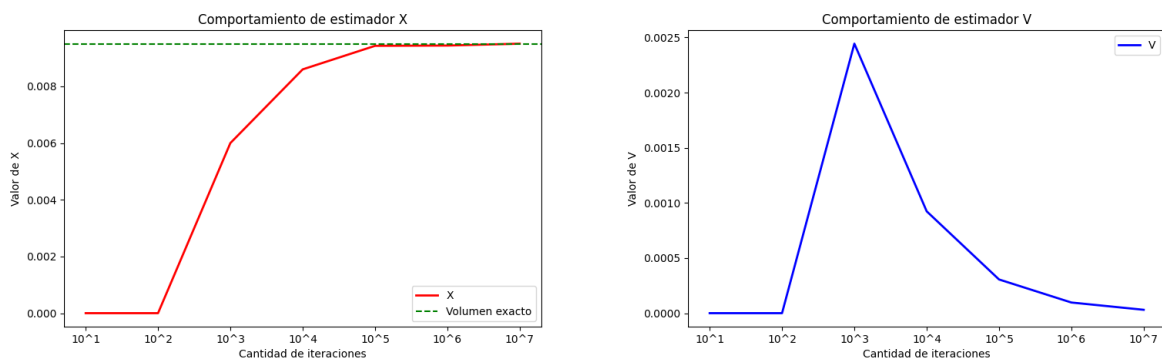


Figura 3.2: Comportamiento de estimadores en cálculo de volumen SIN restricciones

Partiendo de que el valor del volumen "exacto" de la hipersfera es 0,009499628763439042 (siguiendo la fórmula $\pi^3 r^6 / 6$ e ignorando el error numérico de *Python*), se puede comprobar, tanto en la tabla como en las gráficas, como el valor del estimador del volumen converge hacia el valor exacto previamente calculado (representado en la gráfica con la línea verde punteada). Por otra parte, el valor de la varianza disminuye potencialmente convergiendo a 0 como es de esperarse.

Es interesante observar que, en relación al tamaño de muestra y en comparación con los resultados anteriores, el estimador comienza a tener valores distintos a cero para un tamaño de muestra de 10^3 en contraposición con el ejemplo anterior, donde esto sucede con $n = 10^4$. Una potencial razón de este hecho es que la región en este ejemplo es bastante menos acotada, por lo que el algoritmo tiene menor dificultad en encontrar puntos dentro de la región, dado que sólo se trata de la hipersfera.

3.2.2. Ejercicio 4

A continuación, se muestra una tabla comparativa de los resultados obtenidos para cada nivel de confianza $1 - \delta$ y error $\epsilon = 0,01$:

$1 - \delta$ (Nivel de confianza)	n_C (Chebyshev)	n_N (TLC)	n_H (Hoeffding)
99.9 % ($\delta = 0,001$)	2,500,000	27,069	38,005
99.0 % ($\delta = 0,01$)	250,000	16,588	26,492
95.0 % ($\delta = 0,05$)	50,000	9604	18,445

De los resultados obtenidos, se pueden realizar las siguientes observaciones:

- Al disminuir δ , los tamaños de muestra crecen sin importar el criterio utilizado. Esto es coherente, ya que el nivel de confianza aumenta al disminuir el parámetro δ .
- Si bien en este ejercicio el valor de ϵ se mantuvo constante, puede intuirse que el comportamiento del tamaño de la muestra en relación a él será similar a la relación con δ , ya que en todas las fórmulas el error se encuentra en el denominador.
- La derivación a partir de la desigualdad de Chebyshev es la más restrictiva, indicando un tamaño de muestra sustancialmente mayor al disminuir δ en comparación a los otros métodos.
- Por otra parte, la derivación a partir del teorema central del límite es la menos restrictiva, pero introduce diversas fuentes de error, especialmente para casos donde el valor a estimar se acerca a los extremos 0 o 1.