Universidad de Antioquia
Facultad de Ciencias Exactas y Naturales
Área: Física Computacional
Docente: José David Ruiz Álvarez

Proyecto Métodos de Montecarlo

Alexander Valencia[a], Daniel Henao[a]

[a] *Instituto de Física.*

Resumen

Existen métodos numéricos que por sus características particulares se aplican a la solución de problemas muy complejos en múltiples áreas del conocimiento. Lo que se busca principalmente a través de su implementación en un entorno de programación, es lograr el análisis del comportamiento que tiene un sistema real, es decir, producir una simulación. Los métodos de Montecarlo representan un ejemplo conocido de simulación y se aplican a la solución de algunos problemas que permiten tener una interpretación probabilística, además, son útiles en la optimización de modelos y la integración numérica. En el proyecto propuesto se desarrolló el código para generar muestras de Montecarlo, siguiendo una distribución de Landau y usando el método de χ^2, con el propósito de optimizar el valor de los parámetros que describen los datos.

Palabras claves: Landau, Montecarlo, Metrópolis, muestras aleatorias.

I. INTRODUCCIÓN.

Los métodos de Montecarlo se han utilizado para resolver problemas en áreas como la física, las matemáticas, la ingeniería y la economía. Su mejora sistemática se produjo con el desarrollo de la computación, sin embargo los primeros usos como herramienta de investigación se documentaron durante el desarrollo de la bomba atómica, mediante estudios que sugerían la simulación de problemas probabilísticos de hidrodinámica.

La aplicación de los métodos de Montecarlo puede estar orientada al análisis de situaciones que sean estocásticas o deterministas. Según la cantidad estimada que se define de acuerdo al teorema del límite central, los ejemplos usuales en la aproximación, tanto de integrales como de distribuciones incluyen los métodos de tipo estándar, el de muestreo estratificado y el de muestreo por importancia o ponderado.

En general la estructura básica de los métodos de Montecarlo consiste en reducir la varianza de una distribución de datos. En el caso estándar el método se basa en generar una cantidad limitada de números aleatorios para realizar una aproximación.

En el muestreo por importancia se utilizan funciones de densidad de probabilidad conocidas (PDF), elegidas de forma tal que sea factible su implementación. Esto implica producir una cantidad limitada de muestras aleatorias de la PDF para direccionar puntos aleatorios hacia zonas particulares, donde la varianza de la distribución de datos de un experimento puede ser alta

Se pueden presentar situaciones que impiden realizar un muestreo directo de la PDF, lo cual se suple usando algoritmos de computación como el de Metrópolis-Hastings mediante el cual se ejecuta la secuencia de muestras aleatorias para obtener la aproximación de interés.

II. MARCO TEÓRICO.

Para generar un histograma que se ajuste a un determinado conjunto de datos, dada una PDF que satisface condiciones como la relación de normalización $\int_x dx\ p(x) = 1$, y que tiene la estructura o forma adecuada para el ajuste, entonces se utiliza dicha PDF para generar una cantidad limitada de muestras aleatorias $x_1 \cdots x_k$ de $p(x)$ y ubicar N puntos también aleatorios sobre la región que comprende.

Cuando se dificulta realizar un muestreo directo, se tiene la alternativa de desarrollar una secuencia de muestras aleatorias a través del Algoritmo de Metrópolis para aproximar la distribución, ejecutando básicamente los siguientes pasos:

1. Elegir un primer vector de estado arbitrario ϕ_0.
2. Generar aleatoriamente un vector de estado ϕ'.
3. Calcular la relación $\Delta s = -\ln[p(\phi')/p(\phi_0)]$.
4. Establecer las condiciones para aceptar o rechazar los vectores de estado consecutivos,
   - Si $\Delta s < 0 \Rightarrow \phi' = \phi_1$ se acepta como el primer vector de estado.
   - Si $\Delta s > 0$, se genera aleatoriamente un número adicional r, tal que si $r < p(\phi')/p(\phi_0) \Rightarrow \phi' = \phi_1$, o en el caso contrario se rechaza.

Las iteraciones se repiten desde el segundo paso, de modo que el algoritmo intente rastrear todo el espacio muestral, hasta tener ϕ_{n+1} estados que perfilan la PDF una cantidad limitada de $n+1$ veces para desarrollar el análisis de los datos.

Las muestras representativas del comportamiento de datos corresponden a un muestreo sistemático de variables aleatorias, en el cual está fundamentado el método de simulación estadística de Montecarlo.

En el caso de análisis asignado, la PDF apropiada es la función de densidad de probabilidad de Landau y está dada por la expresión:

$$p(x) = \frac{1}{\pi c}\int_0^\infty e^{-t}\cos\left(t\left(\frac{x-\mu}{c}\right) + \frac{2t}{\pi}\ln\left(\frac{t}{c}\right)\right)dt$$

Esta distribución no tiene media o varianza definidos, pero se restringen los valores de μ y c al intervalo $(0, \infty)$, donde la PDF está normalizada a la unidad.

El método de minimización χ^2 es clave para obtener los valores óptimos de los parámetros μ y c de la distribución de Landau, porque permite comparar el ajuste de la distribución de datos experimentales con la distribución de muestras aleatorias, definido a partir de ambos histogramas: el que se obtiene usando los datos experimentales y el que se genera con los datos aleatorios.

III. RESULTADOS Y ANÁLISIS

A partir de los archivos binarios de ROOT, se generaron los histogramas del primer y segundo conjunto de datos, los cuales tienen un total de 85000 y 80000 entradas respectivamente, con una cantidad de 200 bins para ambos.

Cualitativamente y de acuerdo a la visualización gráfica de los datos, se eligió un rango horizontal que comprende valores entre $[0, 900]$ y $[0, 700]$ para obtener los histogramas de los datos reales que se aprecian en las figuras 1 y 2, con un ajuste dado por la PDF de Landau integrada en ROOT y los valores estimados de los parámetros apropiados alrededor de los cuales variar se muestran en la tabla 1.

Item	Datos Daniel	Datos Alexander
μ	130	220
c	15.1	10.0

Tabla 1. Valor apropiado de los parámetros en la PDF de Landau, que ajustan los datos experimentales.

En cuanto a la media comparada con la información de las gráficas se observa que hay una precisión mayor al 71% y 84% respectivamente. Veremos si usando el algoritmo de Metrópolis y el criterio χ^2 obtenemos los mismos valores o cercanos.

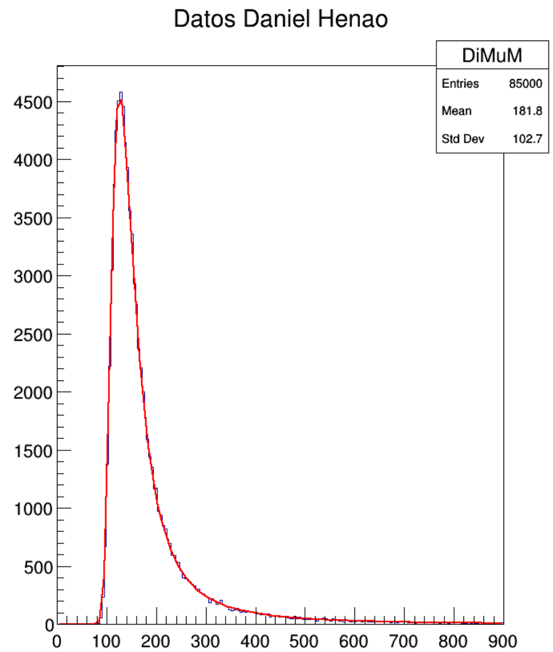

Fig 1. Histograma primer conjunto de datos.

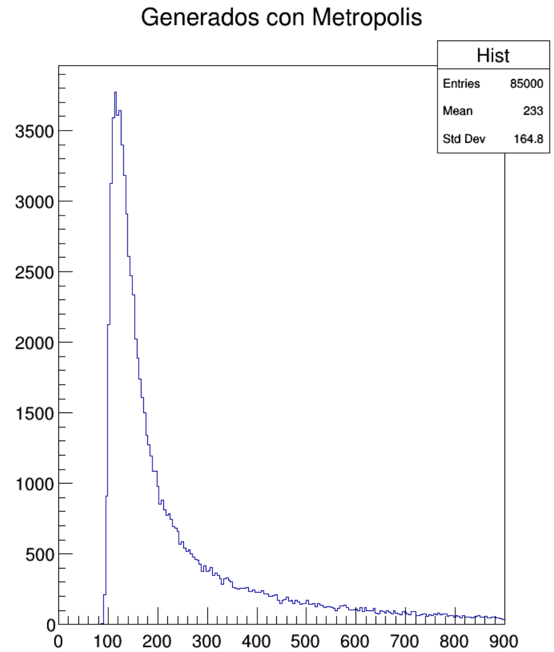

Fig 3. Histograma Metrópolis primer conjunto de datos.

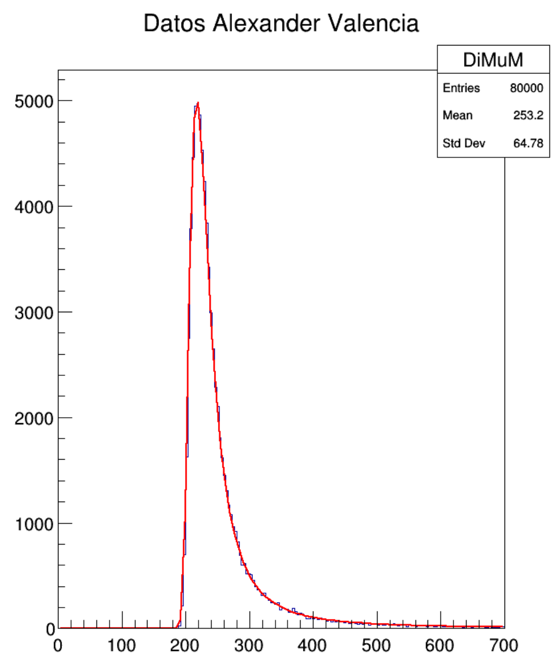

Fig 2. Histograma segundo conjunto de datos.

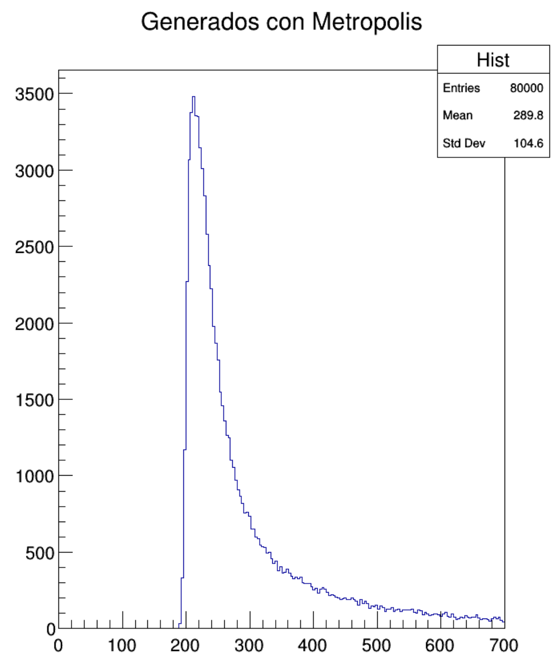

Fig 4. Histograma Metrópolis segundo conjunto de datos.

Usando el algoritmo de Metrópolis y siguiendo la PDF de Landau se generó un conjunto de muestras combinando pares de valores en determinado rango para los parámetros μ y c. Cada conjunto de muestras se comparó con los datos experimentales y se calculó el valor de χ^2 usando las frecuencias de los histogramas. De aquí se encontraron los parámetros correspondientes al menor valor de χ^2, que de acuerdo al criterio de la prueba corresponden a los valores óptimos para la descripción de las distribuciones aleatorias. Los rangos de los parámetros fueron refinándose según la región en que fuera menor, para no tener que hacer un número muy grande de iteraciones de una sola vez. Por esto y porque usamos una PDF predefinida en ROOT, se redujo el tiempo de computo.

En las figuras 3 y 4 se aprecian los histogramas generados a partir de las muestras aleatorias con los mejores parámetros y un pico que se reduce un poco de tamaño comparado con el pico de los datos experimentales, puesto que los histogramas con los datos aleatorios tienen una anchura más pronunciada y esto permite que haya puntos ocupando regiones hacia ambos extremos de la PDF. Pese a esto, se continúa notando la forma característica de la distribución de Landau en los histogramas, lo que indica que el algoritmo se ha implementado de forma apropiada.

Los mejores parámetros se resumen en la siguiente tabla:

Item	Datos Daniel	Datos Alexander
μ	118	214
c	8.6	7.0
χ^2	13455.38	14287.17

Tabla 2. Valor mínimo de χ^2 y valor óptimo de los parámetros en la PDF de Landau, que describen las muestras aleatorias.

IV. CONCLUSIONES

- En situaciones prácticas el método de muestreo por importancia de Montecarlo aplicando el Algoritmo de Metrópolis, es muy útil para lograr una mayor precisión de los resultados. En física por ejemplo, la distribución de Landau está relacionada con el movimiento de partículas cargadas en un medio, por lo tanto la generación aleatoria de datos a través de esta PDF para generar histogramas, permite desarrollar el análisis teórico de las fluctuaciones y el balance de energías, como producto de la interacción entre partículas.

- Debido a que los histogramas generados se ensancharon respecto a los experimentales, los valores de μ y c cambiaron respecto a los ajustes realizados en las figuras 1 y 2. Aun así, son valores cercanos y los menores χ^2, fueron buenos teniendo en cuenta la cantidad de datos, por lo que comparando los datos experimentales con las muestras generadas con el algoritmo de Metrópolis podemos afirmar que, en efecto, los primeros siguen una distribución de Landau y lo hacen con parámetros cercanos a los obtenidos en la Tabla 1.

REFERENCIAS.

[1] ArXiv, repositorio de borradores de artículos académicos (2020). *Introduction to Monte Carlo methods*, [en línea]. Disponible en: https://arxiv.org/pdf/hep-ph/0006269.pdf. [2020, 26 de julio].

[2] WIKIPEDIA, (2020). Método de Montecarlo, Distribución de Landau y Algoritmo de Metropolis-Hastings [en línea]. Disponible en: https://es.wikipedia.org/. [2020, 25 de julio].

[3] UNIVERSITY OF HELSINKY, (2020). *Selected Topics on Data Analysis in Particle Physics*, [en línea]. Disponible en: https://www.mv.helsinki.fi/home/karimaki/CMHEP/Lectures/statLectures.pdf. [2020, 28 de julio].