Análisis de simulación de trayectorias: Caminanta Aleatoria como modelo ideal del movimiento Browniano

Ángela Morales Zamudio
Claudeth Clarissa Hernández Álvarez

Mayo 2019

Abstract

Simulando el problema del caminante borracho o random walk, se hará un análisis estadístico de 100 partículas, trazando trayectorias de 10,000 pasos cada una. El camino trazado por cada partícula fue determinado por un programa generador de números aleatorios, creando así una primera aproximación a un modelo que describa el movimiento Browniano de una partícula.

1 Introducción

El término random walk fue propuesto por Karl Pearson en 1905, quien presentó un modelo sencillo que describía la infestación de mosquitos en un bosque. El modelo describía que en cada intervalo de tiempo regular, el mosquito se mueve una longitud fija a en un ángulo aleatorio. Con esto, Pearson pretendía conocer la distribución de mosquitos luego de cierto tiempo. Después Louis Bachelier, En su tesis doctoral *La Théorie de la Spéculation* propuso a las caminatas aleatorias como el modelo fundamental para las series temporales en el mercado financiero, el ejemplo más habitual que se presenta en la actualidad. El problema del caminante o borracho es, en sí, uno de los casos más sencillos de la modelación de trayectorias de partículas.

2 Marco Teórico

2.1 Caminatas Aleatorias (Random Walks)

Una caminata aleatoria se define como la trayectoria de una partícula en la que su posición en cierto momento sólo depende de su posición en algún instante previo y alguna variable aleatoria que determina su subsecuente dirección y la longitud de paso. En otras palabras, la probabilidad de todos los pasos posibles

estén igualmente distribuídos.
Sean X_k un conjunto de variables aleatorias idénticamente distribuídas (Distribución Uniforme, donde todos los datos tienen la misma probabilidad de ocurrir). La caminata aleaotria S_n se define como:

$$S_n = X_1 + X_2 + ... + X_n = \sum_{i=0}^{n} X_i$$

Donde el valor medio y la varianza de una caminata aleatoria están dados por:

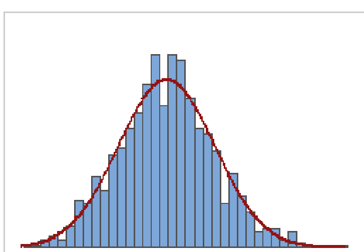
$$\bar{S_n} = \sum_{i=0}^{n} \bar{X}_i \tag{1}$$

$$var(S_n) = \sum_{i}^{n} var(X_i) + 2\sum_{i=i}^{n}\sum_{j\neq i}^{n} cov(X_i, X_j) \tag{2}$$

Esto indica que el comportamiento de los valores medios de las caminatas aleatorias depende de las relaciones entre sus variables aleatorias. Es decir, el comportamiento de la trayectoria va a depender únicamente de cómo se realicen los pasos.
El caso más sencillo de caminata aleatoria (llamada caminata libre) ocurre cuando se tiene la misma probabilidad de avanzar que de retroceder (en cualquier eje). Esto hace que la posición del caminante también tenga valor medio nulo en cada instante.
Desde un punto de vista un poco menos formal, se está diciendo que al medir, uno observa aproximadamente la misma cantidad de pasos en una dirección y en la opuesta.
El ejemplo clásico de este tipo de movimiento es el que observó R. Brown, para una partícula de polen sobre agua [Brown, 1829]. En este caso el movimiento errático de la partícula de polen se debe a los choques que recibe ésta por las partículas del medio que la rodea.

2.2 Teorema del Límite Central

El Teorema del Límite Central dice que la distribución de probabilidad que gobierna $\bar{Y_n}$ se aproxima a una distribución Normal conforme $n \to \infty$, sin importar la distribución que gobierna a las variables aleatorias en cuestión Y_i. Más aún, la media de $\bar{Y_n}$ se aproxima a μ y la desviación estándar a $\frac{\sigma}{\sqrt{n}}$.



3 Metodología

Para el análisis de datos se utilizó el lenguaje de programación Python, empleando las bibliotecas necesarias, entre ellas *scipy.stats*, para poder visualizar gráficas y aplicar pruebas a un conjunto de datos, los cuales fueron generados por medio de simulaciones en $ForTran$ bajo condiciones de aleatoriedad. Gracias a este programa generador de números aleatorios, se logró obtener un conjunto de datos con posiciones X y Y para cada partícula.

3.1 Generador de Números Aleatorios

Ahora se hablará de la manera en que se decició la probabilidad de tomar un paso a la derecha o izquierda (en las posiciones en X), o un paso hacia arriba o abajo (en las posiciones en Y) que se utilizó. Los archivos utilizados para ello contienen dos columnas, cada una con una secuencia de números aleatorios con distribución uniforme. En base a estos números aleatorios con distribución uniforme, se decidieron las probabilidades (0.5 a cada lado, tanto en X y Y). Debido a que la distribución es uniforme, cada vez que se decidió un paso de la partícula, se llamó a un número aleatorio NAX para su movimiento en X, y otro número aleatorio NAY para su movimiento en Y.

En X se toman valores entre 0 y 1.

- Si NAX < 0.5, se desplaza a la derecha.
- Si NAX > 0.5, se desplaza a la izquierda.

En Y se toman valores entre 0.5 y 1.

- Si NAY < 0.75, se desplaza hacia arriba.
- Si NAY > 0.75, se desplaza hacia abbajo.

Esto debe ser consistente con el Teorema del Límite Central, ya que la probabilidad de que la partícula, después de una gran cantidad de pasos, esté a determinados pasos de su origen, debe corresponder a una distribución Gaussiana (Observado en los histogramas anteriores).

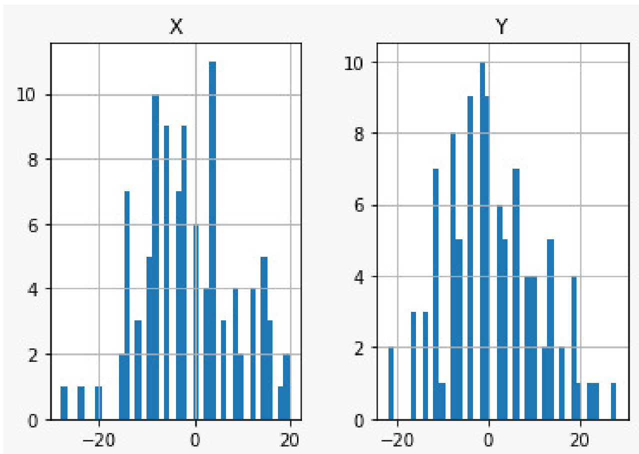
3.2 Distribución de las partículas

Las simulaciones proporcionadas se basaron en un Generador de Números Aleatorios para cada partícula para determinar las parejas (X,Y) de pasos de cada una. Es por ello que se puede decir que, con sencillas simulaciones, se pueden obtener una gran cantidad de datos o archivos de datos de manera aleatoria, cada uno siguiendo su propio "dado de probabilidad" que describe su movimiento.

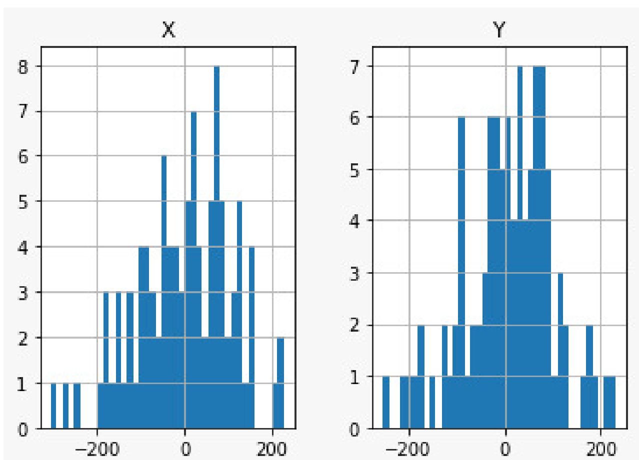
4 Resultados

4.1 Prueba de Hipótesis

Habiendo recolectado en dos DataFrames el paso número $10,000$ y 100 de cada una de las cien partículas, se generaron histogramas para las posiciones en X y en Y, obteniendo los siguientes diagramas:



Histogramas de las posiciones en X y en Y del paso número 100 de las 100 partículas



Histogramas de las posiciones en X y en Y del último paso de las 100 partículas

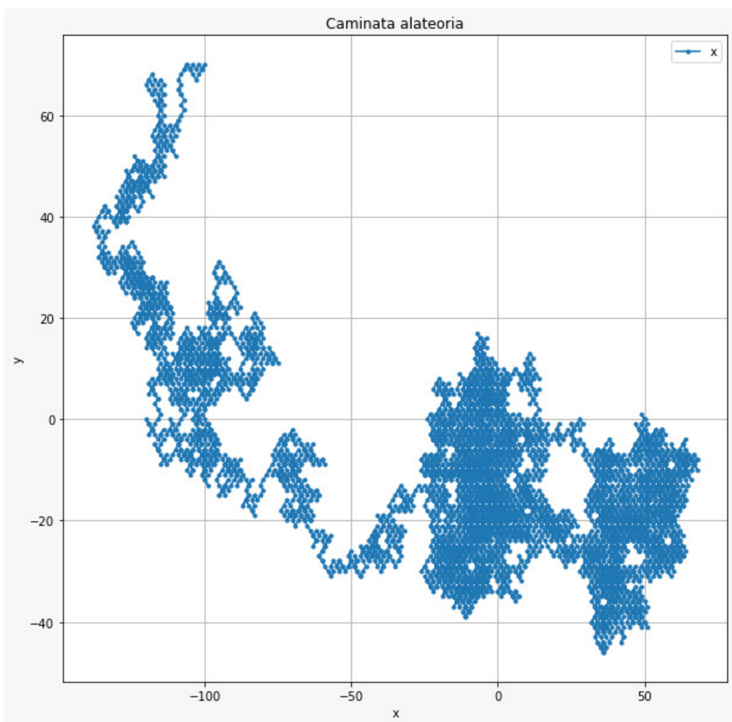
A simple vista se puede observar que parecen tener una distribución Gaussiana (Normal). Usando una prueba de Hipótesis *Shapiro-Wilk*, dado que no se conoce en primera instancia la distribución de estos datos, se probará si siguen dicha distribución.

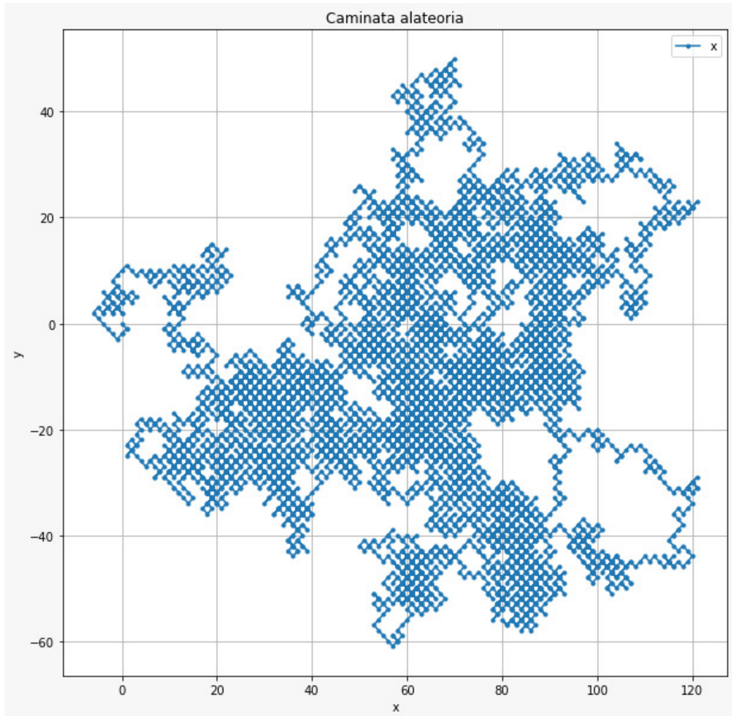
- H_0 = La muestra sigue una Distribución Gaussiana.
- H_1 = La muestra no sigue una Distribución Gaussiana.

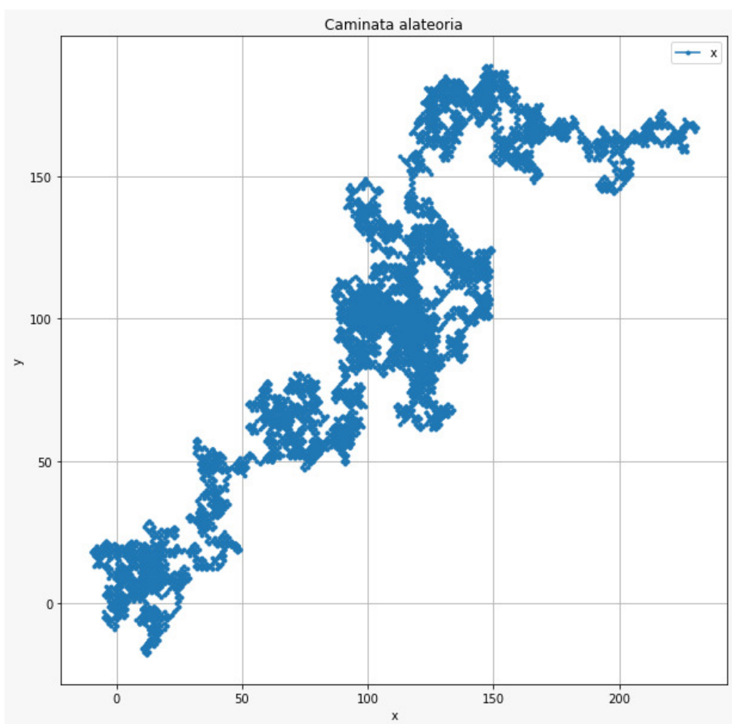
Haciendo uso de las herramientas de Python, en ambas pruebas el p-valor calculado era mayor que el parámetro α, por lo que no se rechazaba la Hipótesis Nula, y se concluía que ambos DataFrames (del paso 100 y del paso 10,000) seguían una distribución Normal, por lo que aquí se puede observar el Teorema del Límite Central.

4.2 Gráficas

Se tomaron tres partículas al azar (de las 100 que teníamos registradas) y se graficaron sus trayectorias (posiciones X,Y).




Caminata alateoria
x
y
x
40
20
0
−20
−40
−60
0
20
40
60
80
100
120


Caminata alateoria
x
y
x
150
100
50
0
0
50
100
150
200

5 Discusión

Una vez analizados los datos de las partículas con caminata aleatoria, puede luego agregarse otras condiciones en las simulaciones que modifiquen el comportamiento de las partículas, ya sea que un lado esté más favorecido que el otro al punto de poder darle un sentido físico más acorde a la realidad como lo que sucede en el movimiento Browniano con un potencial de tipo Lennard Jones.

6 Bibliografía