

Análisis estadístico y modelado numérico de trayectorias de partícula única: mecanismos de difusión y confinamiento

Augusto A. Kielbowicz Valarezo

Tesis de Licenciatura en Ciencias Físicas
Facultad de Ciencias Exactas y Naturales
Universidad de Buenos Aires

Agosto 2017

Tema: Análisis estadístico y modelado numérico de trayectorias de partícula única: mecanismos de difusión y confinamiento.

Alumno: Augusto A. Kielbowicz Valarezo

L.U. N°: 738/11

Lugar de Trabajo: Laboratorio de Electrónica Cuántica,
Departamento de Física, Facultad de Ciencias Exactas y Naturales, Universidad de
Buenos Aires

Directoras del Trabajo: Dra. Laura Estrada , Dra. Graciana Puentes

Fecha de Iniciación: Abril 2016

Fecha de Finalización: Agosto 2017

Fecha de examen: 1 de Septiembre de 2017

Informe Final Aprobado por:

Autor

Directora

Directora

Profesor de
Tesis de Licenciatura

Jurado

Jurado

Jurado

Resumen

En este trabajo se estudiaron las propiedades estadísticas de caminatas aleatorias individuales mediante simulaciones numéricas. Se estudió cómo generar caminatas con distintas condiciones (caminatas libres y con deriva, caminatas confinadas y caminatas dirigidas) y cuales son las características de ensamble distintivas para cada tipo de movimiento, para luego aplicarlas al análisis de trayectorias experimentales obtenidas por métodos de seguimiento de partícula única. Las trayectorias experimentales analizadas sugieren que más de un tipo de movimiento está presente en una misma trayectoria. Fue entonces objetivo de este trabajo definir criterios estadísticos para evaluar si estos comportamientos son reales o son artefactos del movimiento aleatorio.

Se concluye que es posible obtener una descripción cuantitativa los procesos que gobiernan a la caminata a partir de analizar la distribución de los pasos de la caminata. El análisis se realiza considerando los pasos como una variable aleatoria a partir de la cual se obtiene la velocidad de deriva v del valor medio de los pasos, y el coeficiente de difusión D de la varianza de los mismos. Adicionalmente se obtienen correcciones al coeficiente de difusión cuando el movimiento es dirigido a utilizando para esto el ángulo de giro entre dos pasos consecutivos.

Finalmente presentamos el conjunto de herramientas desarrolladas a lo largo del trabajo junto a un protocolo detallado que permite analizar caminatas individuales.

*Non voglio dimostrare niente,
voglio mostrare.*
Federico Fellini

Índice general

1. Introducción y motivaciones	1
2. Resultados numéricos	7
2.1. Caminatas Aleatorias	7
2.1.1. Caminatas libres	8
2.1.2. Caminatas con deriva	10
2.1.3. Caminatas en d dimensiones	12
2.1.4. Caminatas dirigidas	14
2.1.5. Caminatas confinadas	19
2.2. Desplazamiento cuadrático medio (MSD)	21
2.2.1. MSD para movimiento libre	22
2.2.2. Valores del coeficiente de difusión local	27
2.2.3. MSD para movimiento con deriva	27
2.2.4. MSD para movimiento dirigido	30
2.2.5. MSD para movimiento confinado.	31
3. Criterios de segmentación	33
3.1. Ángulo de giro	33
3.2. Probabilidad de escape	36
3.3. Forma de la caminata	40
3.4. Criterio de selección	44
3.4.1. Metodología de análisis	45
4. Análisis de trayectorias experimentales	49
4.1. Dispositivo experimental	49
4.2. Trayectorias experimentales	50

4.3. Análisis de trayectorias experimentales	51
5. Discusión y conclusiones	61
A. Apéndices	65
A.1. Introducción de la Teoría de probabilidad	65
A.1.1. Procesos aleatorios	65
A.1.2. Introducción a la probabilidad	66
A.1.3. Variable aleatoria	67
A.1.4. Operaciones con variables aleatorias	70
A.2. Caminatas Aleatorias	71
A.2.1. Caminatas libres	72
A.2.2. Caminatas con deriva	77
A.2.3. Caminatas en d dimensiones	79
A.2.4. Caminatas confinadas	81
A.2.5. Caminatas dirigidas	81
A.2.6. Ecuación del telégrafo	86
A.2.7. Distribución de probabilidad para caminatas cortas	87
A.2.8. Caminatas de paso fijo (Pearson Walks)	89
A.2.9. Caminatas de paso variable	90
B. Códigos computacionales	91
B.1. <code>randomWalk()</code>	92
B.1.1. Propiedades	92
B.1.2. Métodos disponible en la clase	92

1

Introducción y motivaciones

En las últimas décadas muchas técnicas basadas en la detección de partículas individuales se han convertido en herramientas de uso cotidiano en disciplinas que van desde la física y química, hasta la biología y la nanotecnología [Kusumi et al., 2014, Roberts et al., 2007, Ruhnow et al., 2011]. Las técnicas de partículas individuales permiten determinar distribuciones de propiedades específicas que muchas veces son difíciles de obtener en mediciones de conjunto (*experimentos bulk*) debido al proceso de promediado inherente a las mediciones de conjunto, lo cual puede conducir a resultados erróneos. Por ejemplo, en el interior celular, las partículas pueden interactuar con diversos componentes celulares resultando en poblaciones con diferentes propiedades de movilidad. Adicionalmente, la dinámica de cada población podría cambiar en el tiempo y/o espacio, efecto que también estaría enmascarado en un experimento de conjunto.

Entre las técnicas basadas en la detección de partículas individuales más comúnmente utilizadas, se encuentran los métodos para el seguimiento de partículas individuales (o SPT por sus siglas en inglés). Los métodos de SPT se han constituido en una herramienta muy poderosa para el estudio de la dinámica de procesos que abarcan un amplio rango de escalas temporales y espaciales. En el área de la biología, métodos de SPT fueron aplicados por primera vez durante las décadas del 80 y 90 [Othmer et al., 1988, Qian et al., 1991, Kusumi et al., 1993, Saxton and Jacobson, 1997]. A partir de ese momento la cantidad de aplicaciones ha crecido significativamente principalmente con el avance de las técnicas de microscopía y el desarrollo de marcadores moleculares cada vez más eficientes. Las trayectorias de partículas individuales tienen codificada información tanto del mecanismo del movimiento como de las condiciones del entorno en donde se producen. Sin embargo, la utilización de estas técnicas para responder preguntas específicas está principalmente limitada por dos efectos. Por un lado, la resolución espacial y temporal con la que se puede determinar la posición del marcador; y por el otro, el tiempo total de experimentación que está dado por la durabilidad del marcador fluorescente. Estas dos características determinan el conjunto de preguntas que pueden abordarse.

El marcado celular a partir de moléculas orgánicas, proteínas fluorescentes o puntos cuánticos (*Quantum dots*) es una herramienta que desde hace más de 10 años es extensamente utilizada en microscopías de fluorescencia. Sin embargo, este tipo de marcadores sufren, muchas veces, de procesos fotofísicos y/o fotoquímicos indeseados como el parpadeo (*blinking*) o la degradación fototérmica (*photobleaching*), lo que limita su utilización en experimentos de SPT a solo unos pocos segundos de medición. Conocer la trayectoria de una partícula única por unos pocos segundos puede ser muy revelador en algunos casos y completamente infructuoso en otros. En este sentido, recientemente las nanopartículas (NPs) metálicas empezaron a utilizarse como marcadores alternativos a los tradicionales compuestos fluorescentes en experimentos de microscopía óptica [Durr et al., 2007]. Una de las principales ventajas en el uso de NPs metálicas es el prácticamente infinito tiempo de duración del experimento (decenas de minutos) consecuencia del no desgaste del marcador. Esto que a primera vista es una ventaja, puede llevar a resultados erróneos si la trayectoria no se analiza adecuadamente. Al estar la partícula en un medio heterogéneo como lo es una célula, las partículas pueden sufrir distintos comportamientos, producto de su interacción con los diferentes componentes del interior celular. A modo de ejemplo, en un trabajo muy reciente de Maucort y colaboradores mostraron que la dinámica

1. Introducción y motivaciones

de vesículas¹ en células neuroendocrinas² oscila entre períodos con diferente difusividad o entre períodos de movimiento difusivo y transporte activo [Maucort et al., 2014]. Es claro entonces que en un análisis que promedie la información a lo largo de toda la trayectoria llevará a resultados no representativos de la realidad.

A modo de ejemplo en la figura 1.1 mostramos tres mediciones experimentales, gentileza de la Lic. Manuela Gabriel, de la trayectoria de nanopartículas de oro de 10 nm de diámetro (AuNP) difundiendo dentro del núcleo de células de ovario de hámster chino (CHO-K1). Es evidente que el movimiento de las NPs está gobernado por al menos dos comportamientos diferentes como se desprende de las trayectorias que muestran regiones donde las NP pasan más tiempo (recuadros rojos) y regiones que recorren en menos tiempo (recuadros verdes)

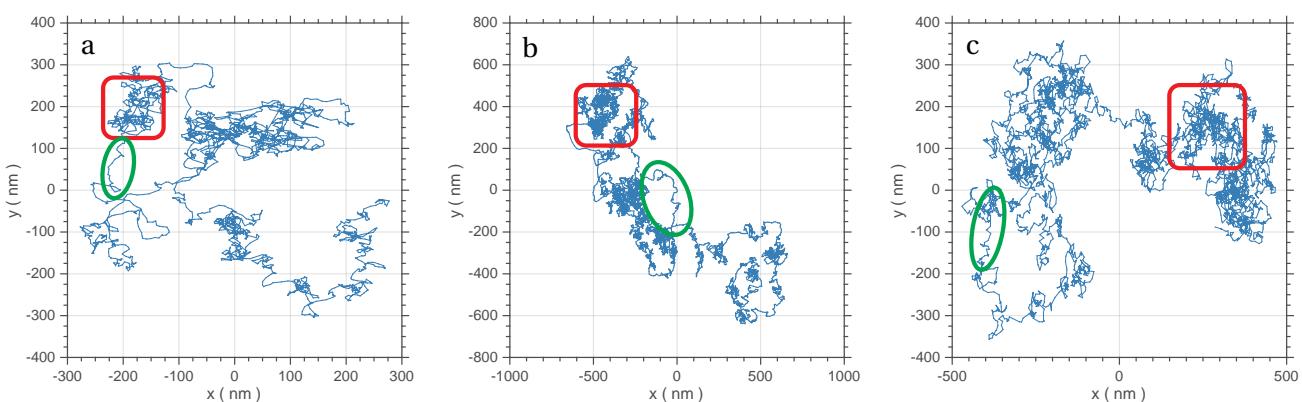


Figura 1.1: Trayectorias de nanopartículas de oro difundiendo en el interior del núcleo de células de ovario de hámster chino (CHO-K1). La duración de los experimentos fue de a) 45 min, b) 38 min y c) 50 min. Encuadradas en rojo las regiones donde la partícula pasa más tiempo y en enmarcadas por elipses verdes las regiones que recorre en menos tiempo.

Análisis de trayectorias de partículas individuales

Para sobrelevar los problemas de mediciones de ensamble, se desarrollaron técnicas enfocadas en el SPT para seguir la posición de partículas en el tiempo [Kusumi et al., 1993, Saxton and Jacobson, 1997]. Dado que la resolución espacial y temporal del método es adecuada, estas trayectorias pueden ser analizadas para extraer información cuantitativa del mecanismo involucrado en el movimiento de la partícula. Ya que cada partícula es observada de manera independiente, SPT puede distinguir fácilmente poblaciones de partículas con diferentes propiedades de movimiento.

Las trayectorias de las partículas dan muestra del mecanismo responsable del movimiento en cada etapa. Esto hace de SPT una técnica importante para entender las dinámicas intracelulares. Sin embargo, el enfoque de utilizar moléculas individuales tiene sus limitaciones. La brevedad de los tiempos de observación y la posibilidad de que las moléculas identificadas no sean representativas de la población, lo cual causa errores de muestreo [Bressloff, 2014].

En su trabajo sobre las aplicaciones del SPT al estudio de la membrana celular Saxton presenta varias preguntas que dejan en claro las capacidades de la metodología [Saxton and Jacobson, 1997]:

¹En biología celular, una vesícula es una estructura que se encuentra en el interior de una célula. Consiste de un compartimiento pequeño y cerrado, separado del citoplasma por una capa lipídica igual que la membrana celular(Wikipedia).

²Las células neuroendocrinas son células que reciben señales neuronales (neurotransmisores liberados por células nerviosas o células neurosecretores) y, como consecuencia de estas señales, liberan moléculas mensaje (hormonas) a la sangre .

- ¿Cómo hacen las partículas para moverse? ¿Y en qué proporción, el movimiento de varias partículas se aleja de la difusión pura? ¿Cómo está controlado ese movimiento y cuál es su función?
- ¿Cómo está organizada la superficie de la célula? En qué proporción las membranas se comportan distinto al modelo de *fluid mosaic*? ¿Es el modelo de tiempo fractal una descripción útil de la superficie celular? ¿Cómo están constituidas las estructuras en la superficie de las células? ¿La compartimentación previene la intercomunicación (*crosstalk*) de los receptores? ¿Qué control regional o global de la dinámica sobre la membrana celular existe?
- ¿Cuáles son los efectos del movimiento heterogéneo en medios heterogéneos sobre la cinemática y el equilibrio(termodinámico)?

Las preguntas ordenadas en tres grupos permiten diferenciar tres características de la información que se puede obtener del análisis de las trayectorias. El primer grupo referido al movimiento propio de las partículas, el segundo grupo toma en consideración al medio en el que se encuentra la partícula y finalmente, cómo las mediciones pueden indicar desviaciones en las consideraciones del modelo estadístico y por ende al equilibrio termodinámico.

La complejidad del análisis se hace evidente principalmente cuando estos mecanismos ocurren de forma simultánea, además de que su importancia relativa puede depender de las proteínas y el tipo de célula en estudio. Muchos de estos mecanismos aparecen simultáneamente: obstrucción por proteínas móviles o inmóviles, unión transitoria con especies móviles o inmóviles, confinamiento por los corrales, unión u obstrucción por matriz celular además de interacciones hidrodinámicas.

Además de todo esto, es instructivo para *calibrar* (*o descalibrar*) la intuición propia, el poder generar caminatas aleatorias que permitan estudiar y optimizar las condiciones experimentales. La simulación numérica de caminatas permite observar difusión aparente, movimiento dirigido, confinamiento y transiciones entre diferentes tipos de movilidad.

El objetivo del análisis de SPT es clasificar trayectorias en los distintos modos de movimiento y generar la distribución de las cantidades que caracterizan ese movimiento, como lo son el coeficiente de difusión, la velocidad, el exponente de difusión anómala, el tamaño de los corrales o la probabilidad de escape.

¿Por qué caminatas aleatorias (*Random Walks*)? Además, ¿qué son?

El término *random walk* fue propuesto por Karl Pearson en 1905, en una carta en *Nature* donde presentó un modelo sencillo que describe la infestación de mosquitos en un bosque [Pearson, 1905]. El modelo describe que cada un 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. Lord Rayleigh respondió la carta contando que él ya había resuelto el problema en una forma más general en 1880 en el contexto de la propagación de ondas de sonido en medios heterogéneos³.

Por otro lado, unos años antes, la teoría de las caminatas aleatorias ya había sido desarrollada por 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[Bachelier, 1900], varias décadas antes de que esta idea se convierta en la base de la Finanza Teórica moderna. Algo importante a recalcar en el trabajo de Bachelier, es el de identificar la utilidad de la teoría de probabilidades para estudiar el comportamiento de algo que hasta aquel momento no tenía forma de ser descripto de manera determinista. Además, parece que fue el primero en ver la conexión entre las caminatas aleatorias discretas y la ecuación de difusión.

³Para conocer la historia completa consultar [Hughes, 1996]

1. Introducción y motivaciones

Por otro lado, el mismo año en que Pearson envió la carta, Albert Einstein publicó, aparentemente desconociendo los trabajos de Rayleigh y Bachelier, su trabajo sobre el *movimiento Browniano* [Brown, 1829, Einstein, 1905]. En este trabajo describe el movimiento de una molécula pesada rodeada de un gas, a partir de una caminata aleatoria. El trabajo de Einstein permitió calcular el coeficiente de difusión de la partícula en función de la viscosidad y la temperatura del gas. Esta teoría fue considerablemente generalizada y extendida por Smoluchowski [von Smoluchowski, 1906]. Además, muchas contribuciones importantes se realizaron posteriormente por Fokker, Planck, Burger, Purth, Ornstein, Uhlenbeck, Chandrasekhar, Kramers, Taylor entre otros. Y por el lado puramente matemático varios aspectos de la teoría fueron analizados por Wiener, Kolmogorov, Feller, Lévy, Doob, Foret y Kac.⁴

Cabe recalcar que en la mayoría de estos trabajos se estudian caminatas aleatorias a partir de su distribución de probabilidad, y en su mayoría la formulación se realiza a partir de la ecuación de difusión o como cadenas de Markov, en el caso discreto.

En este trabajo vamos a tener un enfoque distinto en el cuál no nos concentraremos en la distribución de probabilidad, sino nos enfocamos directamente en los valores medios de la caminata aleatoria pensada como una suma de variables aleatorias. Lo realizamos así ya que de esta forma la descripción del modelo se asemeja a las mediciones de las trayectorias y no tenemos que preocuparnos por las sutilezas matemáticas ni los rangos de validez de los modelos continuos [Chandrasekhar, 1943]. Adicionalmente, en el apéndice proporcionamos una pequeña explicación de cómo se relacionan los distintos modelos, junto con una discusión de los problemas que aparecen en el proceso [Kac, 1947].

Los objetivos principales de este trabajo fueron: Identificar los tipos de movimiento aleatorio que podrían corresponderse con las mediciones. Estudiar mediante simulaciones computacionales cuáles son las características principales de cada movimiento, en particular describiendo cómo son generados los movimientos y luego estudiar cuáles son las propiedades estadísticas de los mismos. Para luego con esto conocer cuáles son los parámetros importantes que se pueden extraer de las trayectorias experimentales y que las describen cuantitativamente.

Además, ya que se observan distintos comportamientos dentro de una misma medición, es necesario definir criterios con los cuales se puedan segmentar las trayectorias de manera automática y basados en las propiedades de cada movimiento de forma que no sean lo más generales y en lo posible independientes del usuario.

La tesis está dividida en tres capítulos en los cuales se presentan tanto resultados numéricos como teoría. Y un apéndice en el cual se complementa con ejemplos y desarrollos matemáticos.

El capítulo 2 se conforma de dos bloques. En el primero se introduce el concepto de caminatas aleatorias a tiempo discreto y se presentan resultados obtenidos a partir de simulaciones computacionales, donde se estudió el comportamiento colectivo o de ensamble en los casos de:

- El caminante puede moverse libremente en una región isótropa y homogénea.
- La presencia de deriva (*drift*) en el movimiento.
- Persistencia en la dirección del movimiento.
- Movimiento en regiones confinadas.

En el segundo bloque se estudia el desplazamiento cuadrático medio (MSD) dado que es una de las herramientas más comúnmente utilizada para el análisis de trayectorias individuales. Se estudiaron cuáles son

⁴El mayor éxito de la teoría de las caminatas aleatorias y el movimiento Browniano se debe al impacto que tuvo al dar fuerte evidencia de la existencia de los "átomos" ya que en ese tiempo se creía que la materia era un continuo.

las falencias del método y como remediar algunas de éstas. Un resultado importante de este capítulo por su aplicabilidad en trayectorias experimentales es la validez de los valores de coeficiente de difusión y velocidad de deriva obtenidos a partir del ajuste de las curvas de MSD.

En el capítulo 3 se definen y analizan criterios de selección binaria para distinguir caminatas libres de caminatas ya sean con deriva, confinadas o dirigidas. Para analizar caminatas dirigidas se evalúa el ángulo entre dos pasos consecutivos y se estudia la probabilidad de que este ángulo sea menor a un valor umbral por varios pasos consecutivos. En el caso de caminatas confinadas se evalúa la probabilidad de que el caminante no se aleje más que una distancia límite del punto del que partió durante un tiempo determinado. Estudiamos cómo depende esta probabilidad con el tiempo y la distancia límite para caminatas libres. Y para detectar caminatas con deriva estudiamos la forma de la caminata mediante la anisotropía de las posiciones que conforman la caminata.

A partir de los distintos criterios de selección definimos algoritmos de segmentación que generan los segmentos por distintas metodologías. Para los casos de movimiento dirigido y confinado los segmentos se generan secuencialmente mientras se va recorriendo la caminata por una sola vez. Y para el caso de caminatas con deriva se evalúa la anisotropía de una parte de la caminata, seleccionada por una ventana que la recorre completamente y luego incrementa de tamaño. Con todos los valores se genera un mapa en función de la posición y el tamaño de la ventana, y se analizan las estructuras que aparecen con forma de 'islas'.

En el capítulo 4 se describe brevemente el dispositivo experimental para la medición de las trayectorias analizadas posteriormente en este mismo capítulo (la medición de las trayectorias experimentales no formó parte de este trabajo de tesis). Luego se propone un protocolo para analizar las trayectorias experimentales a partir de las herramientas numéricas estudiadas en los capítulos anteriores. Se presentan las características que hay que analizar para detectar desviaciones del movimiento libre y varias sugerencias a problemas que pueden aparecer en el análisis. Para esto se explican las propiedades y métodos de la clase `randomWalk()` definida en el entorno Matlab desarrollada para el facilitar el análisis junto con varias funciones para visualizar los resultados.

Y para finalizar, en el capítulo 5 están las conclusiones del trabajo.

1. Introducción y motivaciones

2

Resultados numéricos

Las caminatas aleatorias se utilizan para modelar distintos tipos de movimientos erráticos. Éstos pueden tener distintas características como la distancia promedio que recorren por unidad de tiempo, la presencia de una direccionalidad definida, preferencia por mantener la dirección en la que se está moviendo o la restricción de poder moverse en una región limitada del espacio. En este capítulo generamos caminatas aleatorias con estas características y analizamos los parámetros que las definen.

2.1. Caminatas Aleatorias

Una caminata aleatoria se define como la trayectoria en la que cada paso se realiza de manera aleatoria, formalmente:

Definición. *Sea $\{X_k\}$ un conjunto de variables aleatorias, se define la caminata aleatoria S_n como:*

$$S_n = X_1 + X_2 + \cdots + X_n = \sum_{i=1}^n X_i, \quad (2.1)$$

Utilizando resultados para los valores medios de la suma de variables aleatorias, se tiene que el valor medio y la varianza de una caminata aleatoria están dados por (ver A.1.1):

$$\langle S_n \rangle = \sum_{i=1}^n \langle X_i \rangle, \quad (2.2)$$

$$\text{var}(S_n) = \sum_{i=1}^n \text{var}(X_i) + 2 \sum_{i=1}^n \sum_{j \neq i} \text{cov}(X_i, X_j) \quad (2.3)$$

Esto indica que el comportamiento de los valores medios de las caminatas aleatorias depende de las relaciones que hay entre las variables aleatorias que la componen, es decir, el comportamiento de la trayectoria va a depender de cómo se realicen los pasos.

Primero consideraremos las caminatas donde las variables aleatorias son independientes y están idénticamente distribuidas (i.i.d), de forma que tienen el mismo valor medio, la misma varianza y la covarianza es nula para todo $i \neq j$. Entonces

$$\begin{aligned} \langle S_n \rangle &= n \langle X \rangle, \\ \text{var}(S_n) &= n \text{var}(X) \end{aligned} \quad (2.4)$$

Notar que en este desarrollo no se hizo ninguna suposición sobre el espacio en el que se realizan las caminatas, tampoco si la distribución de pasos es discreta o continua. Lo único que estamos considerando es que el proceso es discreto en el tiempo. También es posible el caso de tiempo continuo, pero no es de interés para este trabajo, en el apéndice mencionamos más sobre esto.

Cuando los pasos son independientes, la caminata aleatoria es un proceso markoviano. La posición a un tiempo $n + 1$ depende sólo de la posición en el tiempo n , es decir, es un proceso sin memoria.

$$S_{n+1} = S_n + X_{n+1}, \quad (2.5)$$

Seguidamente presentamos los resultados para los casos en los que los pasos son independientes entre ellos, cuando existe una tendencia en los pasos, cuando el caminante puede moverse en una región confinada. Y mostramos cómo se extienden los resultados a movimientos en varias dimensiones.

2.1.1. Caminatas libres

El caso más sencillo de una caminata aleatoria ocurre cuando los pasos tienen valor medio nulo, es decir, que hay la misma probabilidad de avanzar que de retroceder. Ésto hace que la posición del caminante también tenga valor medio nulo para todo tiempo. Desde un punto de vista un poco más formal, estamos diciendo que al medir uno observa aproximadamente la misma cantidad de pasos en una dirección y en la opuesta.

Esto es un buen modelo para describir el movimiento que hace uno cuando ya pasaron más de diez minutos esperando el colectivo. Empezamos a dar pasos para un lado y para otro pero, a fin de cuentas, no vamos a ningún lado. Pero ese fue un ejemplo un poco inocente para hacer más tangibles las ideas.

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. En la figura 2.1 se pueden observar 4 simulaciones de caminatas aleatorias en las que los pasos tienen longitud fija de 1 nm.¹.

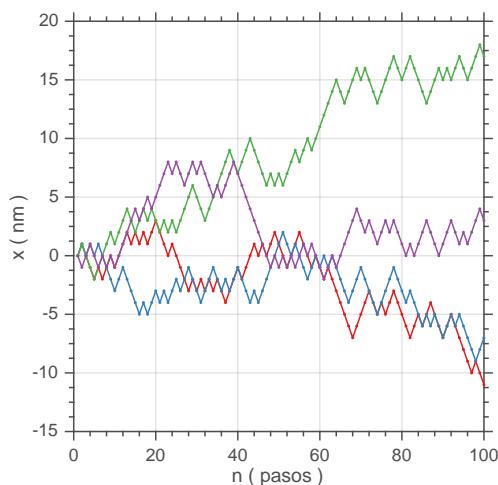


Figura 2.1: Caminatas unidimensionales de 100 pasos realizadas sobre una red. Todos los pasos tienen longitud constante $\ell = 1$ nm.

¹En todo el trabajo vamos a utilizar como unidades de longitud el **nanómetro** (nm), que nos permitirá más adelante unificar las simulaciones con los experimentos

Para ver el comportamiento estadístico de las caminatas aleatorias simulamos 100 mil caminatas de 100 pasos y calculamos el valor medio y la varianza de la posición a cada instante de tiempo. En la figura (2.2a) mostramos el histograma, normalizado a una función de densidad de probabilidad (*pdf*), a distintos tiempos de manera que se puede observar cómo la distribución de las partículas va evolucionando de manera difusiva. Con esto nos referimos a que a medida que avanza el tiempo, la distribución se "aplasta" pero al mismo tiempo se ensancha. Ya que la cantidad total de caminantes es la misma.

Como mencionamos antes, el valor medio de las caminatas es nulo. Ésto se observa en los histogramas, dado que el centro de los mismos se mantiene siempre alrededor de cero. Por otro lado la varianza, o equivalentemente el cuadrado del ancho medio de la distribución, crece de manera lineal con la cantidad de pasos (fig. 2.2b).

El hecho que la varianza crezca de manera lineal implica que la tasa de crecimiento es independiente de la resolución temporal de medición τ que se utiliza. Se define el *Coeficiente de difusión* como [Uhlenbeck and Ornstein, 1930]:

$$D = \frac{var(X)}{2\tau} \quad (2.6)$$

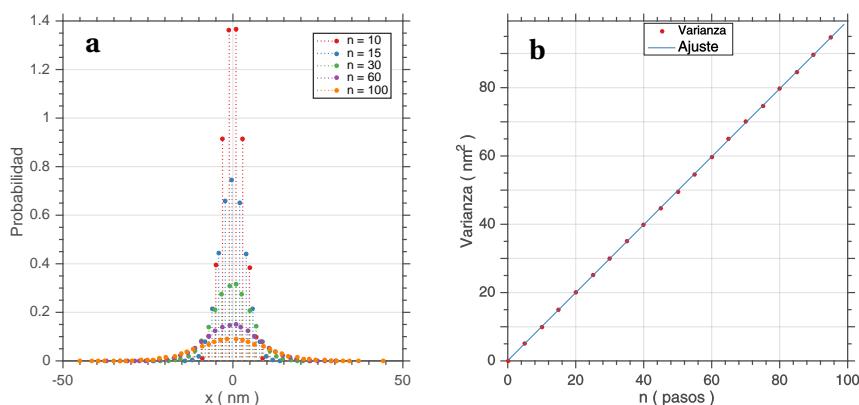


Figura 2.2: (a) Histograma de la probabilidad de encontrar un caminante a una distancia x del origen. (b) Varianza de las posiciones finales después de n pasos. En azul se muestra el ajuste por una recta con $m = 0,990 \pm 0,002$ y $b = 0,09 \pm 0,10$, $R^2 = 1,0$

Con ésto se llega a que el desplazamiento cuadrático medio, luego de un tiempo $t = \tau n$, de un conjunto de partículas que parten desde la misma posición es:

$$\langle x^2(t) \rangle = \langle S_n^2 \rangle = var(S_n) = 2D\tau n = 2Dt. \quad (2.7)$$

En una caminata donde los pasos se realizan de manera independiente, con la misma probabilidad en cada dirección (movimiento es isótropo) y no dependen de la posición en la que está el caminante (espacio homogéneo), el movimiento está caracterizado únicamente por el coeficiente de difusión D el cual es proporcional a la varianza de los pasos.

Hacemos énfasis que este resultado habla de un conjunto o ensamble de partículas, no de una partícula individual. Dado que si uno observa una única partícula, el valor medio de la posición no es nulo. Ésta tiende a alejarse del punto del que partió como se ve en la figura 2.1. Y el desplazamiento cuadrático medio indica

cómo crece la región en el espacio en la que se puede encontrar a la partícula luego de cierto tiempo, no del movimiento propio de la misma.

Antes de concluir con esta sección presentamos un pseudo-código de cómo se realizaron las simulaciones de las caminatas. En el apéndice A.2 presentamos un desarrollo más completo de lo visto hasta ahora, mostramos varios ejemplos de caminatas en las que los pasos son variables aleatorias discretas o continuas y además presentamos brevemente la relación que existe con el caso continuo.

Las simulaciones las hicimos en Matlab aprovechando que es posible vectorizar operaciones². Para generar una caminata a través de un ciclo, en cada paso se genera una valor aleatorio y se lo suma a la posición del paso anterior en la iteración y se guarda el nuevo valor de la posición. Alternativamente, podemos evitar el ciclo vectorizando la generación de los números aleatorios y la suma acumulativa de los mismos. Es decir, en un solo paso nos generamos un vector en el que todos sus elementos son valores aleatorios. Seguido realizamos una suma acumulativa sobre este vector, generando un nuevo vector tal que en cada elemento está la suma de todos los valores del vector original hasta esa posición. El tiempo promedio en generar una caminata de mil pasos es de $t_{loop} = (1,42 \pm 0,08)$ ms utilizando el ciclo y de $t_{vec} = 0,02 \pm 0,01$ ms con el código vectorizado. Además de la mejoría en el tiempo de cómputo esta forma de generar la caminata, nos permite ver que existe una equivalencia entre tener las posiciones de la caminata o los pasos que se dieron.

```

1 nPasos % número de pasos de la caminata
2 longitudPasos % longitud del paso
3 pasos = longitudPasos * random(nPasos) % vector aleatorio de 1 y -1 de longitud n
4 caminata = acumsum(pasos) % suma acumulativa de los pasos

```

2.1.2. Caminatas con deriva

Cuando en el movimiento existe una preferencia en alguna dirección del movimiento, se dice que las caminatas presentan una deriva.

Esto puede ocurrir ya sea porque el medio no es isótropo (por ejemplo si existe un gradiente de concentración en el medio que produce una fuerza constante sobre el caminante) o si hay flujo constante del medio, entre otras posibilidades. Es decir, la variable aleatoria que representa los pasos tiene valor medio no nulo:

$$\langle X \rangle = \ell_0 \neq 0.$$

implicando que las ecuaciones 2.4 toman la forma[Kac, 1947]:

$$\begin{aligned} \langle S_n \rangle &= n \ell_0, \\ var(S_n) &= n var(X) \end{aligned} \tag{2.8}$$

En la figura 2.3 se pueden ver caminatas aleatorias simuladas con pasos generados de una distribución uniforme centrada en $\ell_0 = 0,3$ nm y con una varianza igual a 1 nm. Este movimiento es equivalente a un movimiento libre montado sobre una traslación constante, lo cual se ve claramente en los histogramas de la figura 2.4a, donde se ve como los histogramas mantienen su forma acampanada pero se trasladan, es decir, el valor medio de las posiciones crece.

²Vectorizar una operación quiere decir que se puede realizar una misma operación sobre todos los elementos de un vector sin necesidad de recurrir a una iteración

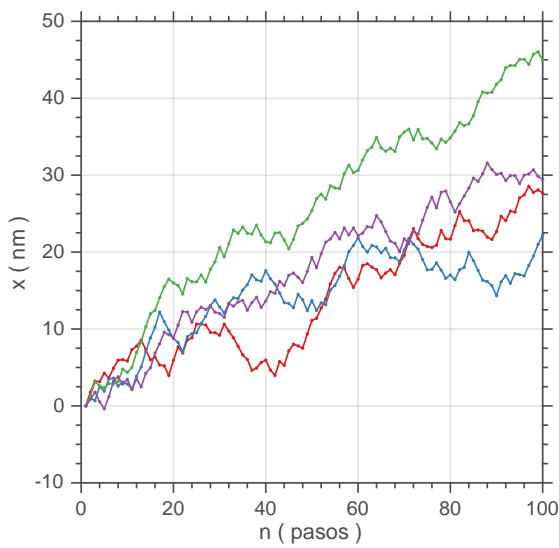


Figura 2.3: Caminatas unidimensionales con distribución de pasos uniformes de varianza $\sigma^2 = 1 \text{ nm}^2$ y valor medio $\ell_0 = 0,3 \text{ nm}$

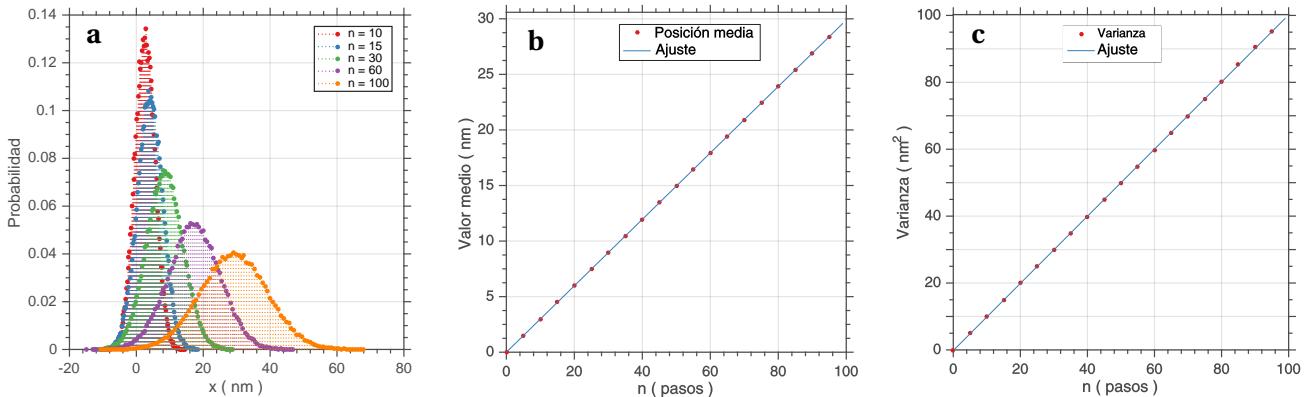


Figura 2.4: (a) Histograma de las distribuciones de probabilidad de encontrar al caminante a una distancia x luego de que dio n pasos.

(b) Valores medios de las posiciones junto con el ajuste (azul) con $m_p = (0,30002 \pm 0,00007) \text{ nm}$, $b_p = (0,001 \pm 0,004) \text{ nm}$.

(c) Varianza de las posiciones finales junto con su ajuste $m_v = (1,002 \pm 0,002) \text{ nm}$, $b_v = (-0,18 \pm 0,14) \text{ nm}$.

Al igual que se definió el coeficiente de difusión en función del tiempo característico τ , definimos la velocidad de deriva como:

$$\nu = \frac{\ell_0}{\tau}. \quad (2.9)$$

Para comprobar que realmente el coeficiente de difusión no depende del valor medio de los pasos, simulamos caminatas con distintos valores medios, calculamos el valor medio y la varianza de la posición a cada tiempo. Ajustamos las curvas por rectas (fig. 2.4b,c) y graficamos las pendientes obtenidas en cada caso como se

muestra en la figura 2.5. La pendiente del valor medio de las posiciones es exactamente el valor medio de los pasos y el coeficiente de difusión no presenta ningún cambio representativo en su valor.

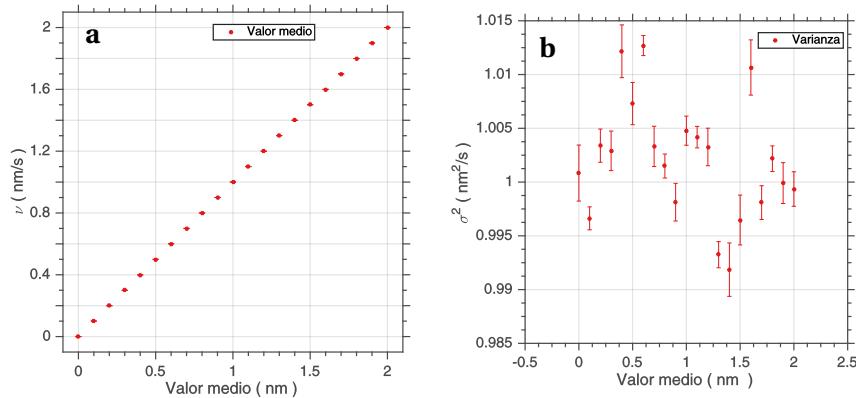


Figura 2.5: Pendiente de los ajustes del valor medio (a) y la varianza (b) de la posición de los caminantes en función del valor medio de los pasos ℓ_0 .

Cuando la distribución de los pasos de una caminata aleatoria tienen valor medio no nulo, las caminatas presentan una deriva. Esta deriva puede ser producto de una tendencia en la dirección que toman los pasos, es decir, tienen distinta probabilidad de ocurrir en cada dirección o también puede estar presente porque la caminata se realiza en un medio que está en movimiento. Una característica importante es la independencia entre la velocidad de deriva y el coeficiente de difusión, recalmando que éste no representa una la velocidad del caminante.

2.1.3. Caminatas en d dimensiones

Considerar una variable aleatoria $\mathbf{X} = [X_{(1)}, X_{(2)}, \dots, X_{(d)}]$ que ahora pertenece a un espacio vectorial de d dimensiones. La distribución de probabilidad de esta variable, $p(\mathbf{X})$ se puede tomar como la distribución conjunta de d variables aleatorias $p(X_{(1)} = x_1, \dots, X_{(d)} = x_d)$, donde cada coordenada es una variable aleatoria $X_{(i)}$.

El movimiento va a depender de las características de esta distribución, es decir, que también depende de las relaciones que existen entre las distintas coordenadas. Por ahora vamos a considerar el caso de movimiento independiente en cada dirección, de manera que $p(\mathbf{X})$ se descompone como el producto de las distribuciones marginales.

En este caso, el valor medio de cada paso como un vector es el valor medio de cada una de sus componentes:

$$\langle \mathbf{X} \rangle = [\langle X_{(1)} \rangle, \langle X_{(2)} \rangle, \dots, \langle X_{(d)} \rangle] \quad (2.10)$$

Y para calcular la varianza, recordamos que $\mathbf{X}^2 = X_{(1)}^2 + X_{(2)}^2 + \dots + X_{(d)}^2 = R^2$, entonces

$$var(\mathbf{X}) = \sum_{i=1}^d var(X_{(i)}) \quad (2.11)$$

Si además los pasos en cada dirección están idénticamente distribuidos (movimiento isotrópico), la varianza

se reduce a:

$$\begin{aligned} \text{var}(\mathbf{X}) &= d \text{var}(X) \\ &= 2dD t, \end{aligned} \quad (2.12)$$

Para generar una caminata aleatoria en d dimensiones, solo hay que modificar el vector que tiene los pasos. En vez de generar un vector de $n\text{Pasos} \times 1$, generamos uno de $n\text{Pasos} \times d$.

- 1 nPasos = % número de pasos de la caminata
- 2 pasos = random(nPasos, dimension) % vector aleatorio de $n \times d$ dimensiones donde cada fila representa un paso
- 3 caminata = acumsum(pasos) % suma acumulativa de pasos

En las figuras 2.6 y 2.7 presentamos ejemplos de caminatas aleatorias simuladas en dos y tres dimensiones.

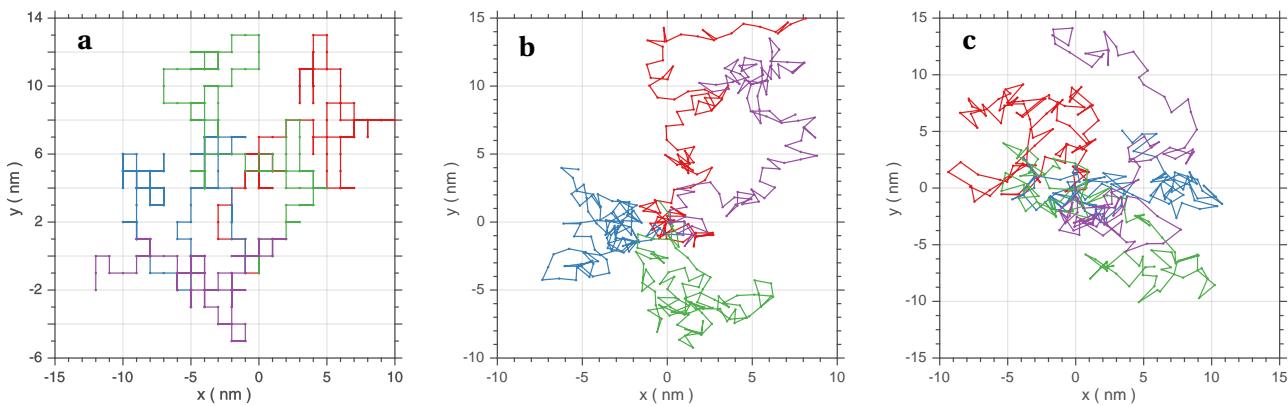


Figura 2.6: Caminatas en dos dimensiones (a) en una red \mathbb{Z}^2 . (b) Caminata con pasos de longitud constante (*Pearson walk*) y (c) caminatas con pasos con distribución uniforme.

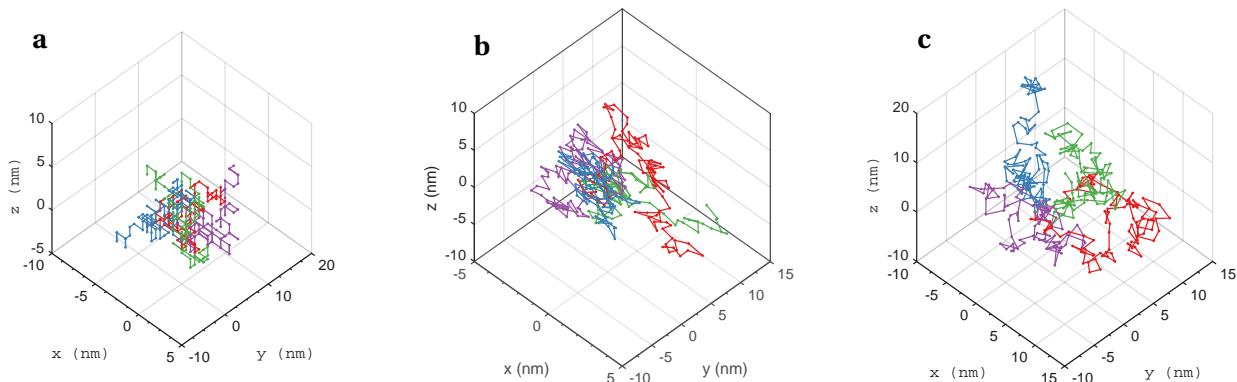


Figura 2.7: Caminatas en tres dimensiones. Las caminatas son generadas en: (a) una red \mathbb{Z}^3 , (b) con longitud de paso constante y (c) con pasos distribuidos de manera gaussiana en cada una de las coordenadas.

Recalcamos que hay que ser cuidadosos y tener claro cómo se define la caminata aleatoria y definir bien el espacio sobre dónde se la realiza. Por ejemplo la caminata que se realiza en la red, se puede considerar como

una caminata en la que el movimiento es independiente en cada dirección si se piensa la caminata dentro de \mathbb{Z}^2 pero si la caminata es en \mathbb{R}^2 el movimiento en las distintas direcciones está correlacionado ya que solo se puede mover en una dirección a la vez. Hacemos notar esto ya que estas consideraciones afectan el resultado final en la interpretación del valor del coeficiente de difusión.

La extensión de caminatas aleatorias a varias dimensiones se logra al utilizar variables aleatorias que pertenecen a espacios vectoriales. En el caso de que el movimiento en cada dimensión sea independiente del movimiento de las demás dimensiones, el movimiento se puede descomponer como caminatas aleatorias independientes en cada una de las coordenadas.

2.1.4. Caminatas dirigidas

Hasta este punto consideramos los casos en los que los pasos son independientes entre ellos. Ahora vamos a analizar que ocurre en el caso contrario.

Primero vamos a considerar caminatas donde los pasos tienen covarianza no nula entre ellos, es decir:

$$\text{cov}(X_i, X_j) \neq 0, \quad (2.13)$$

Supongamos que los pasos tienen una correlación que decresce exponencialmente con el tiempo (número de pasos), renombrando $\text{var}(X) = \sigma^2$:

$$\text{cov}(X_n, X_{n+1}) = \rho \sigma^2 \quad (-1 < \rho < 1). \quad (2.14)$$

donde ρ es la correlación. Siguiendo la derivación detallada en A.2 llegamos a:

$$\text{var}(S_n) = n \sigma^2 \left(\frac{1+\rho}{1-\rho} - \frac{2}{n} \frac{\rho(1-\rho^n)}{(1-\rho)^2} \right), \quad (2.15)$$

En la ecuación 2.15 se puede considerar que los dos términos entre paréntesis describen dos etapas en el movimiento. El primer término está asociado puramente a la difusión y el segundo término se conoce como el término de advección o término balístico. Este último se caracteriza por estar presente solo a tiempos cortos y tienden a 0 cuando $n \rightarrow \infty$. Esto hace que la varianza tienda de manera asintótica a:

$$\text{var}(S_n) = 2 \tilde{D} t, \quad (2.16)$$

donde definimos el coeficiente de difusión efectivo \tilde{D} como:

$$\tilde{D} = \frac{1+\rho}{1-\rho} D_0. \quad \left(D_0 = \frac{\sigma^2}{2\tau} \right) \quad (2.17)$$

No está de más decir que este cociente siempre es positivo. Si la correlación es positiva el coeficiente de difusión efectivo es mayor al del movimiento libre indicando una preferencia a recorrer regiones más amplias y el mismo tiene una divergencia en el límite de la correlación tendiendo a uno, lo que significa que los pasos apuntan siempre en la misma dirección, de manera que no tiene sentido describirlo como un proceso

aleatorio. En el caso en que los pasos estén anticorrelacionados, es decir $\rho < 0$, el coeficiente de difusión efectivo es menor al del movimiento libre, indicando la tendencia de moverse en dirección opuesta con mayor frecuencia.

Es fácil ver que el coeficiente de difusión D crece monótonamente con ρ . Vale $D = 0$ cuando $\rho = -1$ y $D \rightarrow \infty$ cuando $\rho \rightarrow 1$. Para analizar el caso cuando la correlación $\rho \sim 1$, consideremos $\rho = 1 - \varepsilon$, $\varepsilon \rightarrow 0$. Entonces

$$n_c = -\frac{2}{\log \rho} \sim \frac{2}{\varepsilon},$$

$$\frac{1+\rho}{1-\rho} \sim \frac{2}{\varepsilon} = n_c, \quad \frac{2}{n} \frac{\rho(1-\rho^n)}{(1-\rho)^2} \sim -\frac{n_c^2}{2n} \left(\exp \left[-\frac{2n}{n_c} \right] - 1 \right),$$

Llegando a que 2.15 toma la forma

$$\text{var}(S_n) = (\sigma n_c)^2 \left[\frac{n}{n_c} + \frac{1}{2} \left(\exp \left[-\frac{2n}{n_c} \right] - 1 \right) \right], \quad \rho \rightarrow 1. \quad (2.18)$$

En la figura 2.8 se observan caminatas generadas con pasos con correlación positiva, en ellas se ve como existe una tendencia a mantener la dirección de movimiento, generando la impresión de un movimiento dirigido.

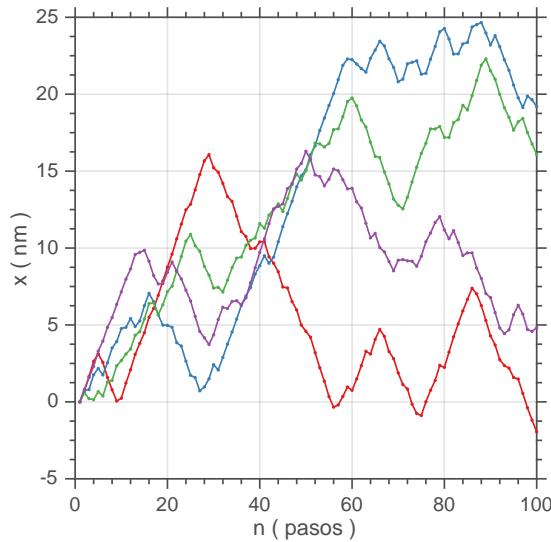


Figura 2.8: Caminatas unidimensionales con movimiento dirigido producto de tener los pasos correlacionados.

Pero si uno observa el movimiento a tiempos más grandes se observa que este comportamiento de persistencia desaparece. Se simularon un conjunto de diez mil caminatas y a partir de las cuales se calculó el valor medio y la varianza de las posiciones a distintos tiempos. En la figura 2.9 se ven los dos comportamientos a distintas escalas temporales, primero un movimiento dirigido y luego difusión libre.

2. Resultados numéricos

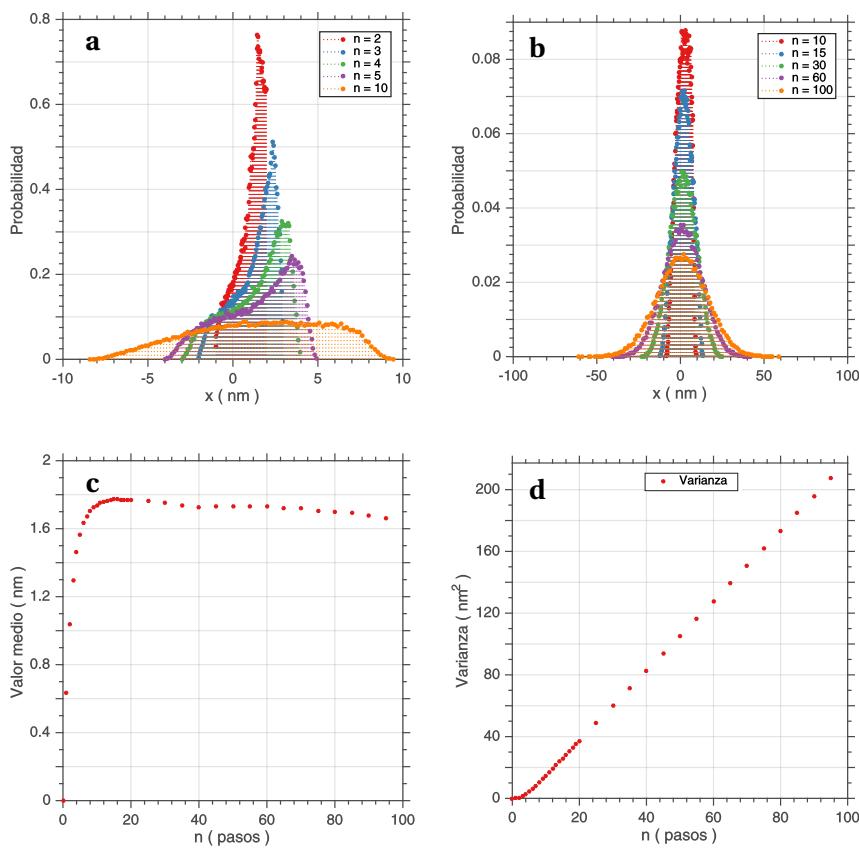


Figura 2.9: (a) y (b) Histogramas de la probabilidad de encontrar al caminante luego de n pasos. Se ve como a tiempos cortos la probabilidad no es simétrica pero a medida que se dan más pasos toman la forma habitual de una campana. (c) Valor medio de la posición de los caminantes y (d) la varianza, en función del número de pasos. Durante los primeros pasos el movimiento tiene una tendencia a avanzar y llega un punto en el que la correlación decae totalmente y el comportamiento es de movimiento libre.

Al observar las distribuciones de la posición de las caminatas, vemos para una cantidad de pasos pequeña $n < 10$ la distribución es asimétrica, de modo que el valor medio no es cero, además inicialmente la varianza crece de manera lenta y luego crece de manera lineal como en el movimiento libre. En la figura 2.10 (varianza en escala logarítmica) se puede ver como el movimiento tiene dos comportamientos.

Como mencionamos en la sección de caminatas en varias dimensiones, uno puede considerar que los movimientos en distintas direcciones están relacionados. Esto no implica que la covarianza entre las direcciones sea nula pero la distribución conjunta no se descompone en el producto de las distribuciones marginales.

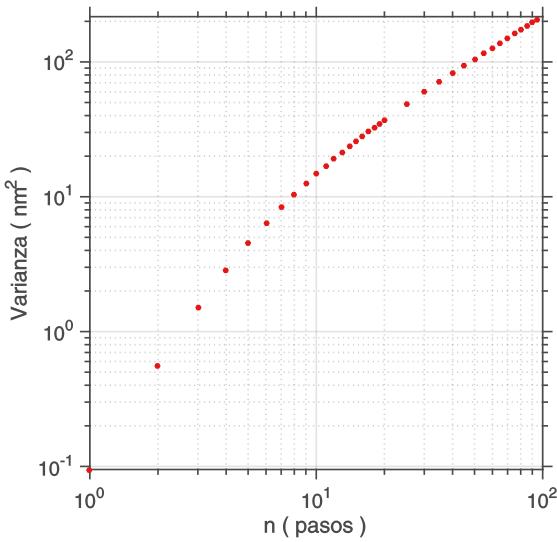


Figura 2.10: Varianza en función del número de pasos. Esta figura es la misma que la figura 2.9 (d) pero en escala logarítmica, de forma de que son evidentes los dos comportamientos.

A continuación analizamos caminatas en dos dimensiones ya que es nuestro caso de interés.

Supongamos que realizamos caminatas donde los pasos son aleatorios pero con la condición de que el ángulo de giro respecto al paso anterior esté acotado en un rango. En un experimento este efecto es observable si se tiene suficiente resolución espacial y temporal para seguir a la partícula. En este caso es posible ver que las mediciones están correlacionadas entre ellas y que el movimiento de la partícula no cambia tan bruscamente su dirección de movimiento sino que tiene cierta inercia.

Si el ángulo de giro de la partícula está acotado entre $[-\pi/2, \pi/2]$, ésta tiene preferencia para moverse en la misma dirección. Esto no restringe a que si observamos la distribución angular de los pasos, ésta no sea uniforme en todo el intervalo $[0, 2\pi]$. Dado que, si la partícula se mueve por suficiente tiempo puede dar una vuelta completa, de manera que apunta en cualquier dirección.

Para analizar este caso bidimensional es útil definir las siguientes variables aleatorias. El ángulo de giro respecto de la m -ésima posición ϕ_m y la longitud del paso l_m . Podemos suponer que estas variables son independientes, de forma que tenemos:

$$\langle l \rangle = \int_0^\infty l p(l) d l, \quad (2.19)$$

$$\langle l^2 \rangle = \int_0^\infty l^2 p(l) d l \quad (2.20)$$

$$c = \langle \cos(\phi) \rangle = \int_{-\pi}^{\pi} \cos(\phi) g(\phi) d \phi, \quad (2.21)$$

$$s = \langle \sin(\phi) \rangle = \int_{-\pi}^{\pi} \sin(\phi) g(\phi) d \phi \quad (2.22)$$

2. Resultados numéricos

donde $g(\phi)$ es la distribución de los ángulos de giro.

Considerando el caso en el que la distribución de ángulos de giro es simétrica, por lo tanto $s = 0$, se llega a [Kareiva and Shigesada, 1983]:

$$var(S_n) = \langle l \rangle^2 \left(n \left(\frac{1+c}{1-c} + b^2 \right) - \frac{2c(1-c^n)}{(1-c)^2} \right). \quad (2.23)$$

La ecuación 2.23 tiene la misma forma que la ecuación 2.15, a menos del término $b^2 = \langle l^2 \rangle / \langle l \rangle^2 - 1$ que es el coeficiente de variación de la longitud de los pasos [Codling et al., 2008]. El cuál es el cociente entre la varianza y el valor medio de la longitud de los pasos y aparece porque se está considerando una variable que sólo toma valores positivos.

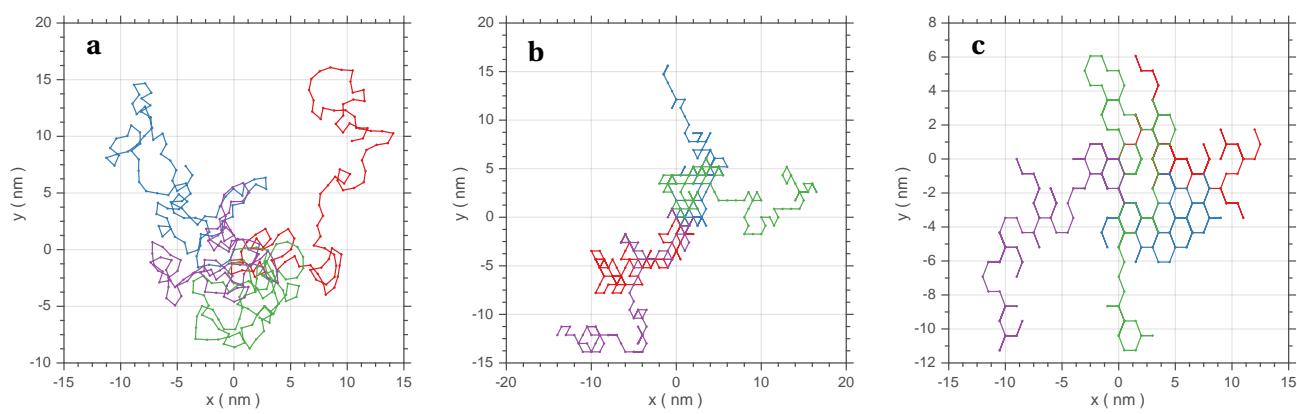


Figura 2.11: Caminatas en dos dimensiones de pasos correlacionados. **(a)** Correlación generada al imponer que los pasos tengan un giro máximo de 90° . **(b)** Los pasos solo pueden dar giros en múltiplos de 60° que da como resultado una red triangular. **(c)** Los pasos solo pueden dar giros en múltiplos impares de 60° que da como resultado una red hexagonal.

```

1 % Código para generar caminatas correlacionadas
2
3 nPasos % número de pasos de la caminata
4 anguloMaximo % ángulo máximo de giro
5
6 longitudesPasos % vector con las longitudes de los pasos
7
8 angulosDeGiro = random(nPasos, [-anguloMaximo, anguloMaximo]) % vector con ángulos de giro
9     generados de manera aleatoria
10 angulosCartesianos = acumsum(angulosDeGiro) % suma acumulativa % Dirección de los pasos respecto
11     del eje x
12
13 pasosNormalizados = [cos(angulosCartesianos), sen(angulosCartesianos)] % vector de nx2 dimensiones , donde
14     cada fila representa el paso
15 pasos = longitudesPasos * pasosNormalizados % Multiplicamos cada paso por una longitud
16
17 caminata = acumsum(pasos) % suma acumulativa de los pasos
  
```

Cuando existe una relación temporal entre los pasos, el movimiento de la caminata aleatoria es distinto en dos escalas temporales. A tiempos menores a un tiempo crítico el movimiento presenta una direcciónalidad específica, donde el valor medio de las posiciones crece. Luego, del tiempo crítico el movimiento se asemeja

a un movimiento libre pero con un coeficiente de difusión efectivo. En el caso bidimensional esta relación puede surgir si se restringe el ángulo máximo que puede girar el caminante entre pasos sucesivos.

2.1.5. Caminatas confinadas

Cuando se empieza a estudiar caminatas en las que el caminante sólo puede recorrer una región acotada del espacio, es más sencillo analizar el sistema utilizando otras herramientas. Considerando el movimiento de la partícula como una cadena de Markov y estudiando cómo es la evolución de la distribución de probabilidad de la posición bajo la condición de: la probabilidad de encontrar a la partícula fuera de la región de confinamiento es nula.

Es posible considerar el problema a tiempo discreto pero el desarrollo es demasiado engorroso y los resultados son complicados de interpretar [Kac, 1947]. Bajo la suposición de que la región de confinamiento tiene el tamaño suficiente para que sea válida la aproximación continua del movimiento libre, se puede resolver el problema a partir de la ecuación de difusión ya que es más sencillo aplicar las condiciones de contorno [Saxton, 1995].

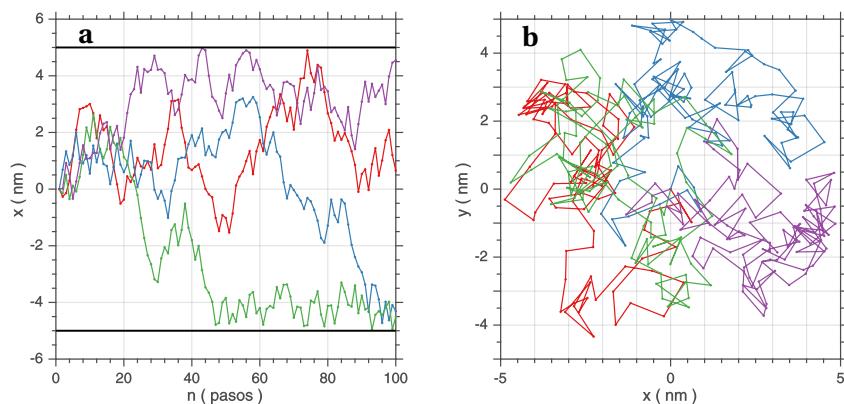


Figura 2.12: Ejemplos de caminatas simuladas, confinadas en (a) una dimensión y (b) dos dimensiones.

En este caso de movimiento confinado es necesario generar la caminata mediante una iteración, debido a que hay que comprobar la posición de la partícula luego de que dé cada paso.

```

1 nPasos % número de pasos de la caminata
2 bordeConfinamiento % borde del confinamiento
3 i = 2 % índice de la iteración. Inicializo en 2 de forma que la caminata comience en el origen.
4
5 caminata = ceros(nPasos, dimension) % inicializar el vector con ceros
6
7 % Si la posición está permitida la agregamos a la caminata, sino,
8 % la descartamos e intentamos otra vez
9
10 while i ≤ nPasos
11
12     paso = random(1,dimension) % paso aleatorio
13     posicion = caminata(i-1) + paso;
14
15     if abs(posicion) ≤ bordeConfinamiento
16         caminata(i) = posicion;
17         i = i + 1;
18     end %if

```

2. Resultados numéricos

19 `end %while`

Tenemos que el valor medio se mantiene nulo si estamos considerando que el confinamiento es simétrico respecto del origen, en caso contrario se va a observar una corrimiento hacia el centro de masa de la región de confinamiento.

Notar que en la figura 2.13a a la distribución le toma tiempo deformarse respecto del perfil acampanado habitual. Esto se debe a que es necesario que el caminante interactúe un par de veces con los bordes del corral para notar su presencia. En el gráfico de la varianza figura 2.13b se ve que al principio la varianza crece de manera lineal como en el movimiento libre y luego se estabiliza y tiende al valor máximo permitido por el confinamiento. Se muestran varias curvas de varianza en función del radio de confinamiento 2.13c. Se ve cómo a pocos pasos todas siguen la misma curva lineal y luego se separan. Junto está el valor máximo que alcanza la varianza en función del radio de confinamiento 2.13d, que crece de manera cuadrática.

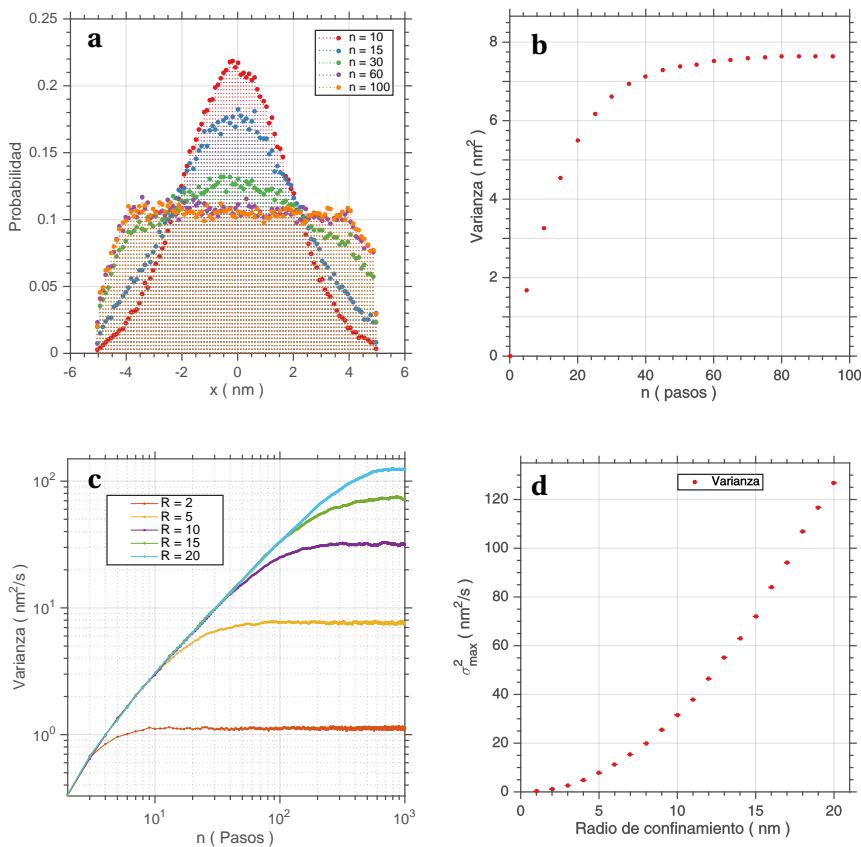


Figura 2.13: (a)Histogramas de las caminatas para distintos tiempos.(b) Valores de la varianza de la posición en función del número de pasos.(c) Curvas de la varianza de la posición para distintos tamaños de confinamiento (d) Valores máximos de la varianza en función del radio de confinamiento.

Haciendo un análisis dimensional se puede considerar que la descripción del movimiento depende de un único parámetro $D t / R^2$ [Saxton, 1995]. A partir de este se puede analizar dos escalas temporales. Primero $t < \sim t_c = R^2 / D$ en la que el caminante aún no interactúa con el confinamiento y, por lo tanto, se mueve libremente y $t > t_c$ donde la partícula recorre varias veces la región y se puede observar el confinamiento.

Resumiendo, hasta ahora analizamos distintos tipos de caminatas aleatorias, definidas como la suma de variables aleatorias idénticamente distribuidas. Analizamos el comportamiento del valor medio $\langle S_n \rangle$ y la va-

rianza $var(S_n)$ de las caminatas en función del tiempo y de los primeros momentos de la distribución de probabilidad de los pasos. Vimos que para caminatas en la que los pasos son independientes, tanto el valor medio y la varianza de las caminatas crece de manera lineal con el tiempo. Y cuando el movimiento es isótropo, es decir hay la misma probabilidad de moverse en una dirección y la opuesta, el valor medio es nulo para todo tiempo.

Cuando los pasos no son independientes, aparecen dos rangos temporales en el movimiento. A escala menores a un valor crítico t_c el valor medio de las caminatas crece, esto se observa como una persistencia en la dirección de movimiento. A tiempos mayores al tiempo crítico, el valor medio se mantiene casi constante y la varianza crece de manera lineal asemejando un movimiento libre pero con un coeficiente de difusión efectivo.

Si el movimiento está confinado, también aparecen dos escalas temporales determinadas por el tiempo crítico $t_c = R^2/D$. Donde las caminatas se comportan como libres a tiempos menores a t_c ya que no interactúan con las paredes del confinamiento.

Un punto importante es que debido a la aleatoriedad de los procesos, ocurre que si uno mira una única caminata puede ver que los distintos tipos de movimientos están presentes a pesar de haber sido generados con uno solo. A modo de ejemplo, en la figura 2.14 mostramos tres caminatas simuladas con distintas características pero en las que aparenta a simple vista haber más de un tipo de movimiento.

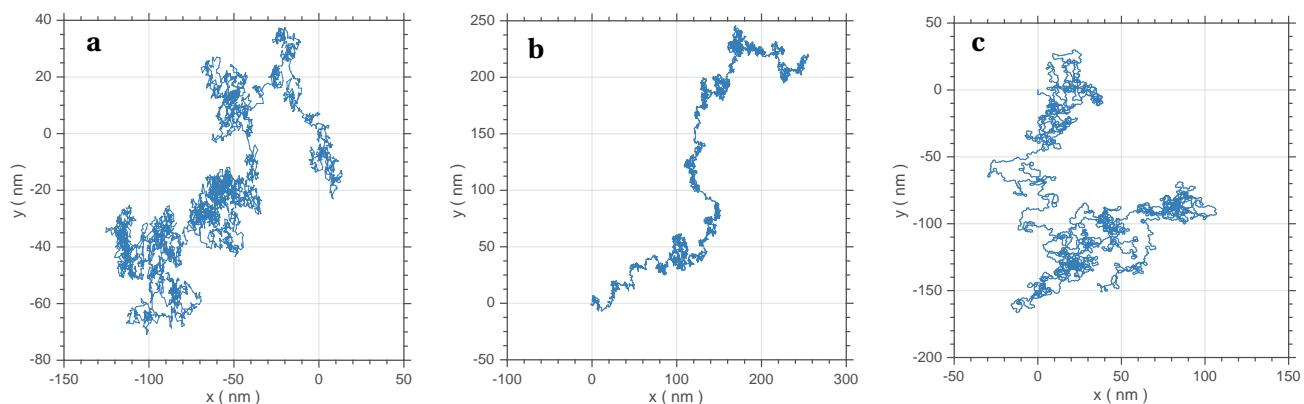


Figura 2.14: Caminatas aleatorias simuladas. Todas las trayectorias tienen 5000 pasos y consisten en (a) caminata gaussiana de pasos independientes en cada dirección, (b) caminata gaussiana con deriva y (c) caminata con correlación debida a que el ángulo máximo de giro de los pasos de 90°

2.2. Desplazamiento cuadrático medio (MSD)

Como vimos en la sección anterior, las caminatas aleatorias de distintos tipos tienen diferentes valores medios y varianzas. Para unificar los resultados considerar que cada tipo de movimiento puede ser descripto por el desplazamiento cuadrático medio (msd) dado que éste es la suma $\langle X^2 \rangle = var(X) + \langle X \rangle^2$. El msd es una herramienta ampliamente utilizada para analizar trayectorias de partículas individuales [Qian et al., 1991, Saxton, 1993a, Kusumi et al., 1993]. Sin embargo ésta tiene sutilezas, debidas al origen estocástico del movimiento, que son muy frecuentemente discutidas en la literatura [Michalet, 2010, Coscoy et al., 2007] a tener en cuenta cuando se realiza el análisis .

Formalmente el valor del msd se calcula a partir de la distribución de probabilidad:

$$\text{msd}(t) = \langle R^2(t) \rangle = \int_0^\infty r^2 p_R(r; t) d r, \quad (2.24)$$

esto es un inconveniente si se quiere analizar caminatas individuales ya que no es posible obtener la distribución de probabilidad para todos los tiempos. Pero en el caso de tener una trayectoria única en la que se puede suponer que el movimiento es ergódico, es decir todos los pasos están idénticamente distribuidos, es equivalente tomar el promedio temporal a tomar el promedio sobre el ensamble [Landau and Lifshitz, 1969].

Para calcular el *msd* a partir del promedio temporal es necesario considerar todos los posibles pares de puntos de la trayectoria que estén distanciados temporalmente por cierto $t = \tau n$ y calcular los valores medios del cuadrado de la diferencia de estos pares. Obteniendo así que el *msd* es:

$$\text{msd}(n) = \frac{1}{N - (n + 1)} \sum_{i=1}^{N-(n+1)} (x_{i+n} - x_i)^2, \quad (2.25)$$

donde los $x_i = x(i\tau) = x(t)$ son las posiciones que fue tomando la partícula, n es lo que se denomina "*tiempo de retraso*" y es la variable de la cual depende el *msd* y N es el número total de pasos de la trayectoria, es decir, su duración.

Expresiones analíticas para los distintos tipos de movimiento pueden hallarse en la literatura [Saxton and Jacobson, 1997, Codling and Hill, 2005]:

$$\text{msd}(t) = \begin{cases} 2dDt & \text{para movimiento libre,} \\ 2dDt + (\nu t)^2 & \text{para movimiento con deriva,} \\ R_0^2 \left(1 - A_1 \exp \left[\frac{-2dA_2 t}{R_0^2} \right] \right) & \text{para movimiento confinado,} \\ 2D \left(t - \frac{1}{2\lambda} \left\{ 1 - \frac{1}{2\lambda} \exp[-2\lambda t] \right\} \right) & \text{para movimiento dirigido,} \\ 2dDt^\gamma & \text{para movimiento anómalo.} \end{cases} \quad (2.26)$$

Para los casos de las caminatas libres, con deriva estas curvas coinciden con los resultados obtenidos en la sección anterior. Para el movimiento dirigido es posible reescribir la ecuación 2.15, obtenida en la sección anterior, de manera que tiene la misma forma funcional de 2.26(ver apéndice A.2). El resultado analítico para el movimiento confinado está calculado en el trabajo de Saxton [Saxton, 1995]. Por completitud presentamos el caso anómalo en el apéndice.

A continuación realizamos un análisis de cada una de las curvas y las consideraciones que hay que tener al analizar las curvas experimentales.

De la misma manera en la que se analizó la posición del caminante como una variable aleatoria, se puede analizar el *msd* para cada tiempo de retraso como una función de la variable aleatoria de la posición de manera que ésta tiene todas las características de una variable aleatoria. Entonces al evaluarse a partir de una muestra se tiene un error que depende del número de puntos o mediciones que se tiene del mismo.

2.2.1. MSD para movimiento libre

Para cada caminata la curva del *msd* va a diferir del valor esperado ya que se está promediando sobre un número finito de puntos. Para estudiar cómo difieren los valores del *msd* del valor esperado hicimos simula-

ciones de caminatas, calculamos el *msd* de cada una y trabajamos sobre los valores estadísticos de éstas.

En la figura 2.15 se ve como las curvas de *msd* pueden diferir bastante de la curva esperada. Si se consideran más puntos, ya sea porque se promedia sobre varias trayectorias (fig. 2.16) o porque se toma una trayectoria de mayor duración, el valor de la dispersión de los puntos es menor.

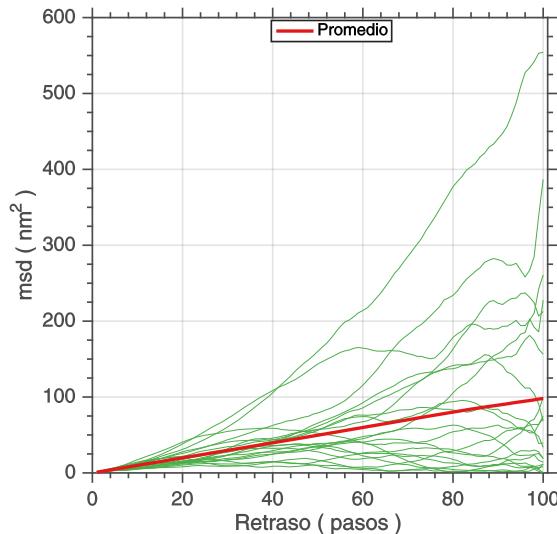


Figura 2.15: Curvas de *msd* de caminatas individuales junto con la curva promedio. Se ve como el comportamiento lineal es característico del promedio de las trayectorias y no corresponde al *MSD* de una única de ellas.

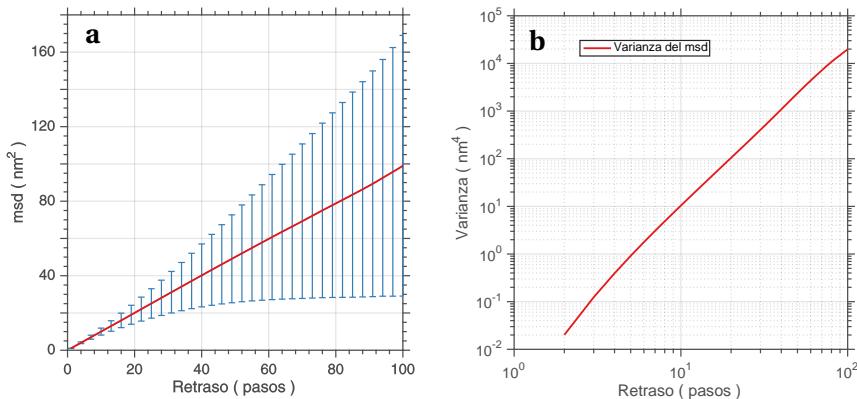


Figura 2.16: (a) Curva del *msd* promedio junto con las barras de error que están dadas por la desviación estándar. (b) Varianza del *msd* en función del número de pasos de retraso.

Kusumi indica que la curva teórica de la varianza del *msd* es $\text{var}(\text{msd}(t)) = (3D t)^2$ [Kusumi et al., 1993]. En nuestro caso obtuvimos a partir de un ajuste del logaritmo de la curva de la figura 2.16b que la varianza crece como una ley de potencia con un exponente de $\delta = 3,112 \pm 0,034$. No encontramos ninguna explicación para esta diferencia, y queda como una duda abierta de este trabajo.

Conocer la varianza de la curva de *msd* es importante si se la quiere ajustar mediante cuadrados mínimos

2. Resultados numéricos

dado que es una hipótesis del método que todos los puntos tengan la misma varianza (*homoscedasticity*). En el caso de no tener la misma varianza (*heteroscedasticity*) se tienen que utilizar cuadrados mínimos pesados (WLS) donde el peso de los puntos va como $1/\sigma^2$ si se conoce la varianza teórica del modelo [NIST, 2012, Bevington and Robinson, 1992].

Boyer estudió el problema de encontrar un estimador óptimo para el coeficiente de difusión mediante cuadrados mínimos a partir de una única caminata aleatoria libre [Boyer et al., 2013]. Analizó el caso de utilizar pesos de la forma $(t - t_o)^{-\alpha}$, obteniendo que el caso de $\alpha = 2$ es el óptimo.

Además del problema de tener que utilizar cuadrados mínimos pesados, surge la pregunta de cuál es la cantidad de puntos del msd que se deben ajustar. Saxton menciona que lo habitual es ajustar solo un cuarto de la cantidad de puntos de la caminata [Saxton and Jacobson, 1997]. Pero ocurre que otra de las hipótesis del modelo de cuadrados mínimos es que los errores tienen que estar distribuidos de manera normal, o a lo sumo de manera simétrica, alrededor del valor esperado. En el caso del *msd* esto no ocurre, ya que tiene una distribución chi-cuadrado en el caso de caminatas con distribución de pasos gaussianas. En la figura 2.17 se muestran los histogramas no normalizados del *msd* para distintos valores temporales. A tiempos de retraso menores a 20 los histogramas se mantienen casi simétricos.

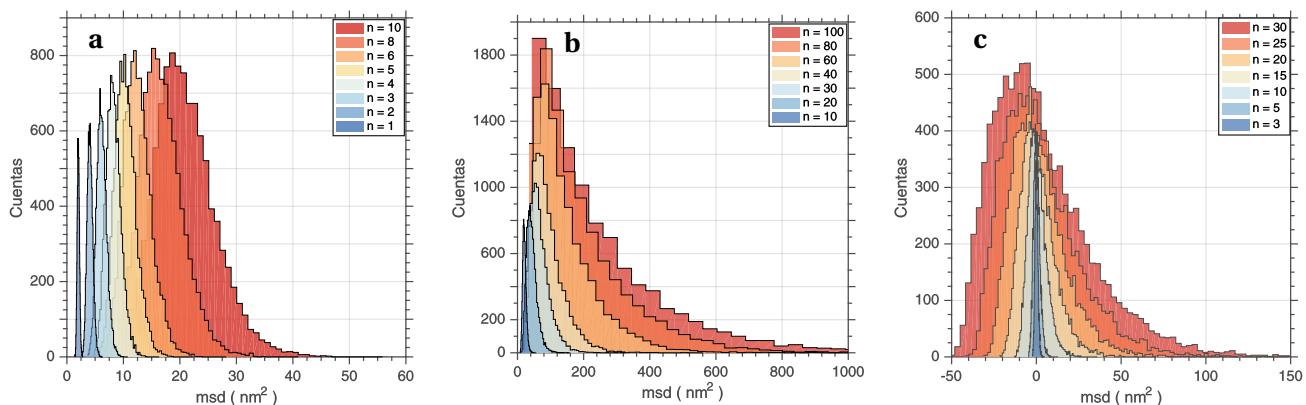


Figura 2.17: Histogramas de los valores del msd simulado. **(a)** Histogramas de tiempos de retrasos menores a 10 pasos **(b)** en tiempos de retraso hasta 100 pasos. **(c)** Histogramas centrados en el valor medio, en este caso se puede observar que a medida de que el tiempo de retraso aumenta, la distribución va perdiendo su simetría respecto del valor medio.

En la figura 2.18 mostramos una curva de msd calculada para una caminata libre de cien pasos. Junto están los ajustes por cuadrados mínimos, ordinarios y pesados, de la misma considerando los primeros 25 puntos. Elegimos este ejemplo ya que a simple vista se puede suponer que la curva pertenece a un movimiento dirigido y eso es un problema al realizar los ajustes experimentales ya que se está imponiendo el modelo por el cuál se quiere ajustar, de manera que puede dar resultados incorrectos si se lo utiliza como criterio para clasificar los distintos tipos de caminatas. En la figura 2.18b se muestran los valores de la pendiente del msd obtenidos ajustando distintos números de puntos, junto con una línea que marca el valor de la varianza de los pasos.

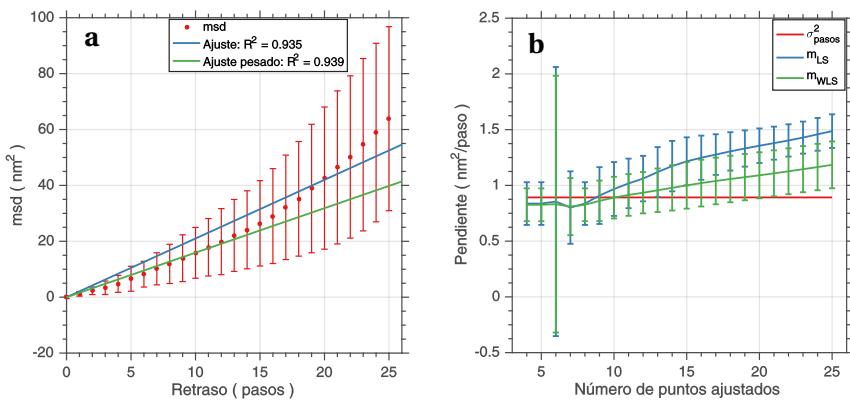


Figura 2.18: (a) Curva de msd junto con los ajustes por cuadrados mínimos y cuadrados mínimos pesados. (b) Valores de la pendiente del ajuste en función del número de puntos ajustados en el msd. Junto con la varianza de la longitud de los pasos σ^2 .

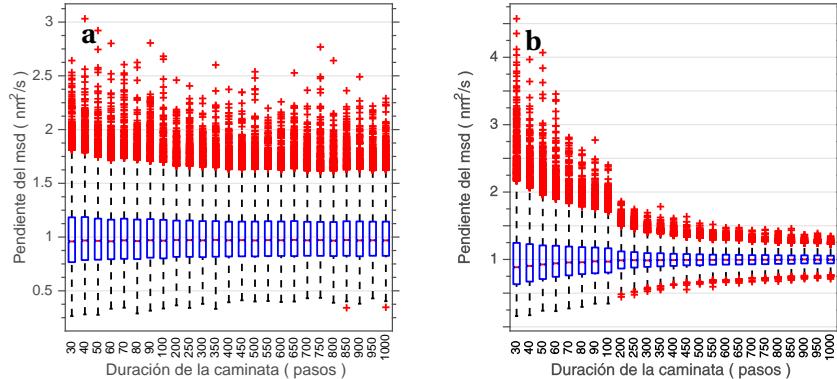


Figura 2.19: (a) Boxplot de los valores de la pendiente del msd obtenidos a partir del ajuste por cuadrados mínimos pesados. Los ajustes se realizaron a tantos puntos como el diez por ciento de la duración de la caminata.(b) También se realizaron ajustes pero ahora se ajustaron los diez primeros puntos del msd independientemente de la duración de la caminata. El box plot indica la mediana con una línea roja, los bordes de la caja azul representan los 25 y 75 percentiles, las barras negras marcan los bordes de la mediana $\pm 2,7\sigma$ de manera que todos los datos más alejados que esta distancia son considerados como *outlier*, cruces rojas. Estos bordes representan el 99.3 si la distribución es normal

2. Resultados numéricos

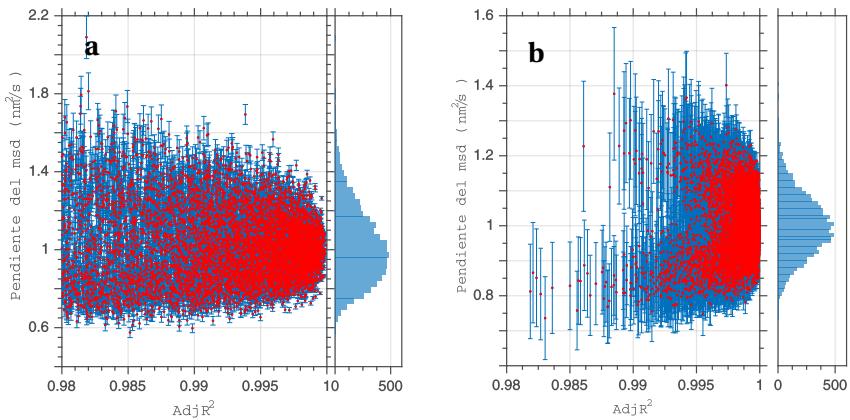


Figura 2.20: Valores de la pendiente del ajuste de curvas de msd de caminatas de mil pasos. Se ajustaron con cuadrados mínimos pesados los diez primeros puntos **(a)** y los cien primeros puntos **(b)**.

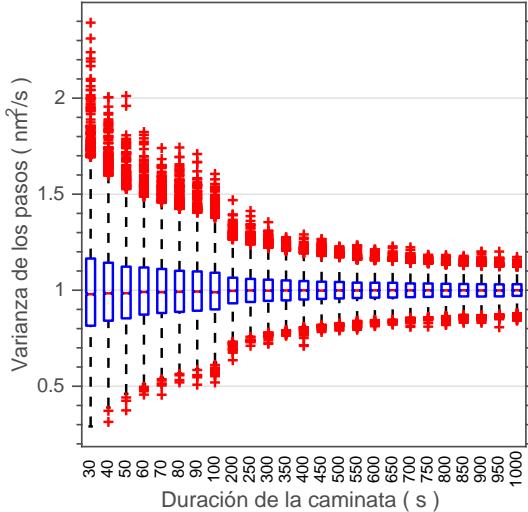


Figura 2.21: Boxplot de los valores de la varianza de los pasos para caminatas simuladas distintas duraciones. Las caminatas se simularon con varianza de los pasos 1 nm^2 .

Otra característica del análisis a tener en cuenta es el modelo por el que se ajusta en el caso de caminatas libre. Se puede ajustar por una lineal estimando la pendiente y la ordenada al origen (ε) o se puede forzar a que la ordenada al origen sea nula. Michalet indica que en el caso de tener error en la medición de la posición es necesario estimar la ordenada al origen a partir del ajuste y que no se debe agregar *a priori* el valor $(0, 0)$ en los datos del msd [Michalet, 2010].

Respecto del mismo problema de tener error en determinación de la posición de la partícula, Martin indica que este error produce un comportamiento anómalo subdifusivo a tiempos cortos [Martin et al., 2002]. Indica que si se observa el msd en escala logarítmica la pendiente aparente a un tiempo de retraso t está dada por:

$$\alpha_{ap} = \frac{1}{1 + 2\varepsilon/4Dt} \quad (2.27)$$

2.2.2. Valores del coeficiente de difusión local

Kusumi define unos coeficientes de difusión locales considerando solo los primeros puntos del msd [Kusumi et al., 1993]. Suponiendo que se tienen suficientes pasos en la caminata para que la varianza de los valores del msd sea despreciable hasta un tiempo de retraso de cuatro pasos, se calculan analíticamente los valores de la pendiente del ajuste a partir de las fórmulas de cuadrados mínimos. Varios términos de las fórmulas se anulan ya que el valor del msd a tiempo de retraso cero es nulo.

$$D_2 = \text{msd}(1) - \text{msd}(0) = v a r(X), \quad (2.28)$$

$$D_3 = \frac{\text{msd}(2) - \text{msd}(0)}{2}, \quad (2.29)$$

$$D_4 = \frac{-3 \text{msd}(0) - \text{msd}(1) + \text{msd}(2) + 3 \text{msd}(3)}{10}, \quad (2.30)$$

$$D_{2-4} = \frac{\text{msd}(4) - \text{msd}(2)}{2}. \quad (2.31)$$

En la figura 2.22 mostramos las distribuciones de estos coeficientes y como están correlacionados entre ellos.

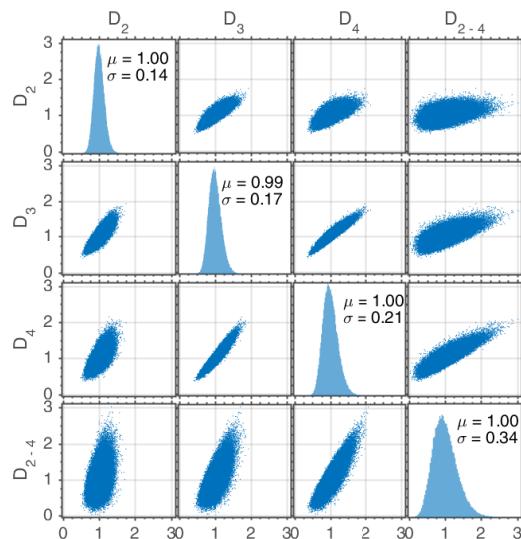


Figura 2.22: Histogramas y gráficos de correlación de los distintos coeficientes de difusión obtenidos a partir de simulaciones de caminatas aleatorias de cien pasos. Los parámetros μ y σ^2 son el valor medio y la varianza de los datos representados en cada histograma.

2.2.3. MSD para movimiento con deriva

La curva teórica del *msd* para movimiento dirigido es una cuadrática sin término constante. Ésta se puede reescribir como $\text{msd}(t) = 2D t (1 + v^2 t / 2D)$, de manera que para observar el efecto de la deriva se tiene que medir por suficiente tiempo como para que el término que tiene la velocidad sea comparable a 1. Calculando la pendiente del *msd* en el origen se tiene que la recta tangente en este punto es exactamente el *msd* del movimiento libre, lo que hace más complicado determinar si existe deriva cuando esta es muy débil.

2. Resultados numéricos

La variabilidad de las curvas individuales del msd sigue estando presente en el caso del movimiento dirigido y dado que ahora se tienen más parámetros de ajuste no realizamos un análisis tan exhaustivo como en el caso anterior.

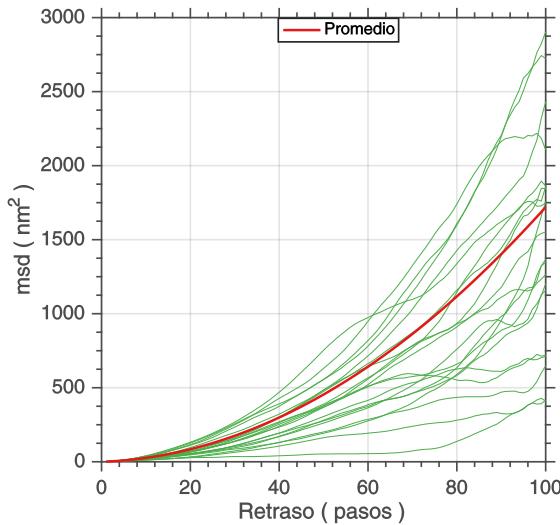


Figura 2.23: Curvas del msd de caminatas con deriva junto con el msd promedio.

En la figura 2.24 mostramos la curva promedio con el error estándar de la misma, se puede observar que el error relativo es mucho menor en este caso.

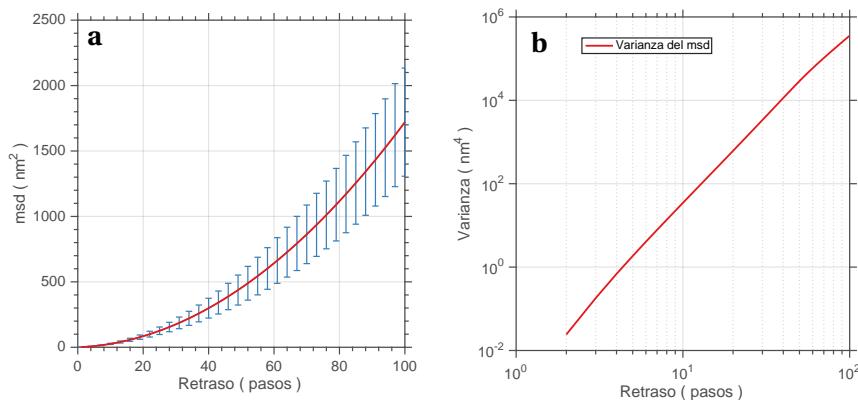


Figura 2.24: (a) Curva del msd promedio, para movimiento con deriva, junto con las barras de error dadas por su desviación estándar. (b) Varianza de los valores del msd en función del tiempo,

Para ver cuán preciso son los valores que se obtienen a partir de ajustes por cuadrados mínimos, ordinarios y pesados, graficamos los valores obtenidos en función del número de puntos utilizados en el ajuste. En la figura 2.25a se muestra la curva del msd junto con los ajustes utilizando los veinticinco puntos. En la figura 2.25b se muestra el coeficiente de correlación de los ajustes en función del número de puntos ajustados. Se ve que los ajustes son adecuados pero si se observa la figura 2.26b la variabilidad del coeficiente del término lineal hace dudar de la efectividad del método para obtener resultado coherente.

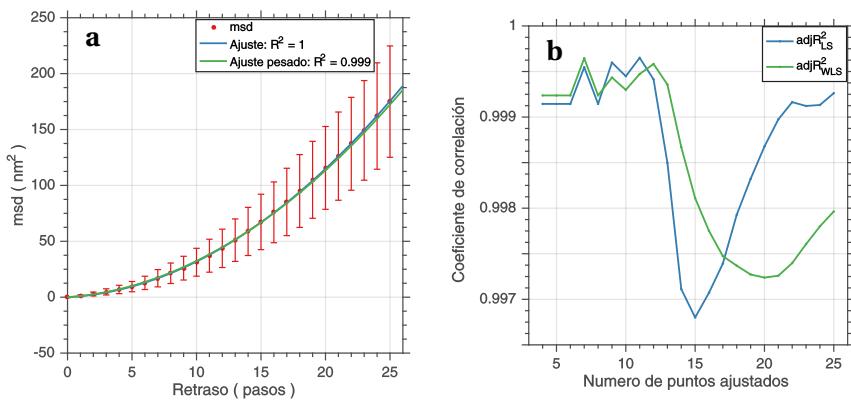


Figura 2.25: (a) Curva del msd de una caminata con deriva junto con los ajustes de cuadrados mínimos lineales pesados y no pesado.(b) Valores del coeficiente de correlación ajustado.

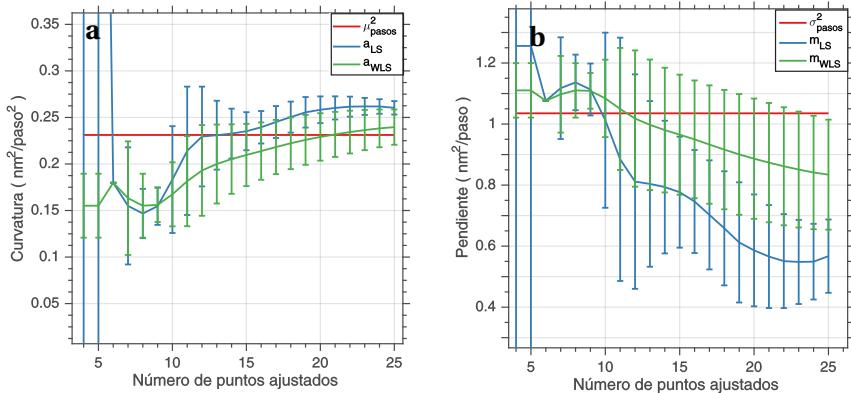


Figura 2.26: (a) Valores de la pendiente del ajuste en función del número de puntos ajustados en el msd. Junto con la varianza de la longitud de los pasos σ^2 .(b) Ordenada al origen del ajuste en función de la cantidad de puntos utilizados para el ajuste junto con el valor medio de los pasos μ .

Nuevamente, los resultados teóricos sólo se asemejan en el caso de tener un ensamble de caminatas. En la figura 2.27 mostramos valores medios de los parámetros de ajuste con sus errores estándar en función del valor medio de los pasos con los que se simularon las caminatas.

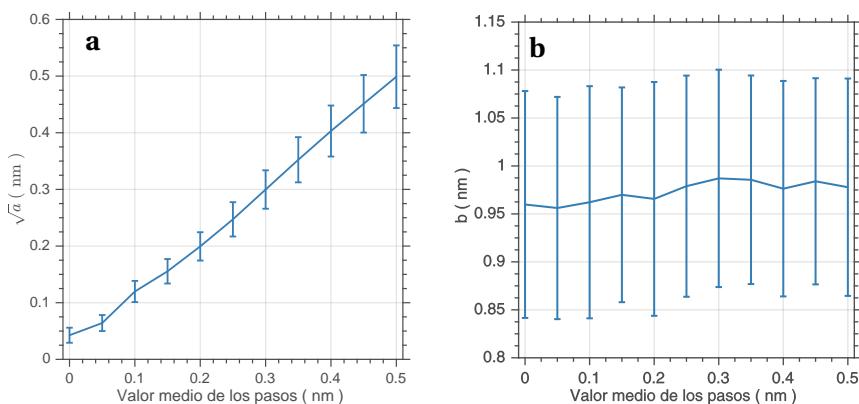


Figura 2.27: (a) Raíz del coeficiente cuadrático del ajuste del msd, de caminatas con deriva, en función del valor medio de los pasos con los que se genera la caminata.(b) Valores del coeficiente del término lineal del ajuste del msd en función del valor medio de los pasos

2.2.4. MSD para movimiento dirigido

Para el caso del movimiento dirigido, la forma funcional del msd tiene dependencias no lineales con los parámetros por lo que no nos preocupamos por estudiar los ajustes. Simplemente presentamos gráficos de del msd para distintas caminatas individuales de 100 pasos junto con la curva promedio 2.28 y la curva promedio con la su error estándar 2.28a y la varianza del msd 2.28b la cual muestra una tendencia muy diferente a la de los casos anteriores.

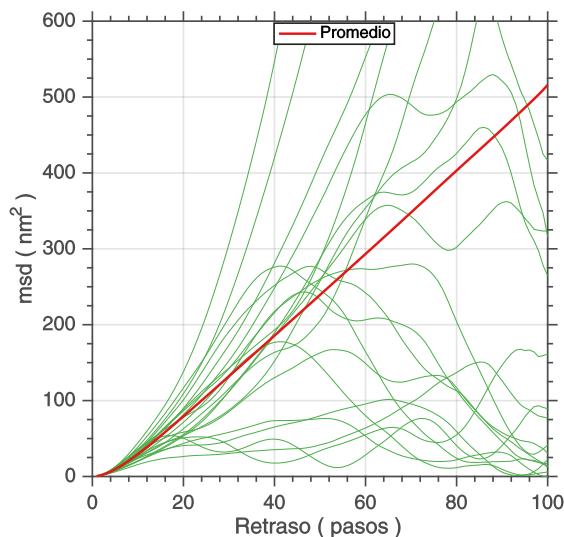


Figura 2.28: Curvas del msd de caminatas con correlación junto con la curva promedio.

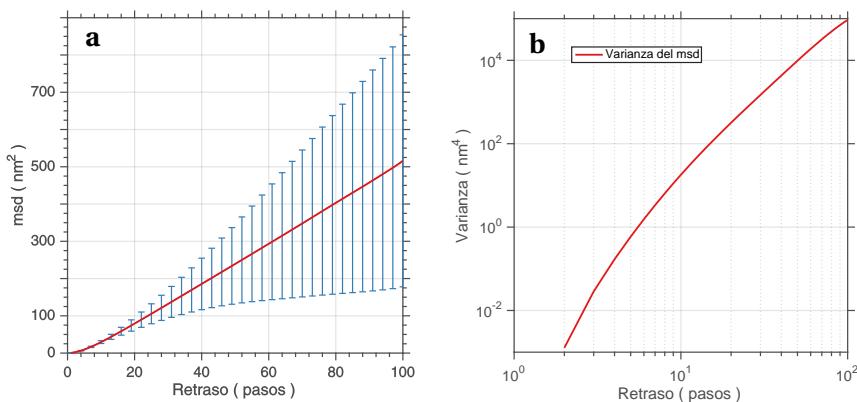


Figura 2.29: (a) Curva del msd promedio junto con sus barras de error para caminatas con movimiento correlacionado.(b) Varianza del msd

2.2.5. MSD para movimiento confinado.

El movimiento confinado presenta dependencias no lineales con los parámetros por lo tanto no analizamos los ajustes de las curvas individuales.

Para generar las figuras simulamos caminatas aleatorias de cien pasos. Los pasos estaban distribuidos de manera normal con valor medio nulo y varianza 1 nm². Las caminatas se realizaron en una dimensión y con un confinamiento de 5 nm.

El valor al que tiende la curva es el radio de giro de toda la región de confinamiento. Esto es así dado que si la caminata tiene suficiente tiempo para recorrer toda la región accesible la varianza de las posiciones de la caminata va a ser equivalente al momento de inercia de un cuerpo con densidad uniforme con la forma del confinamiento. Por lo tanto este valor depende de la forma del confinamiento.

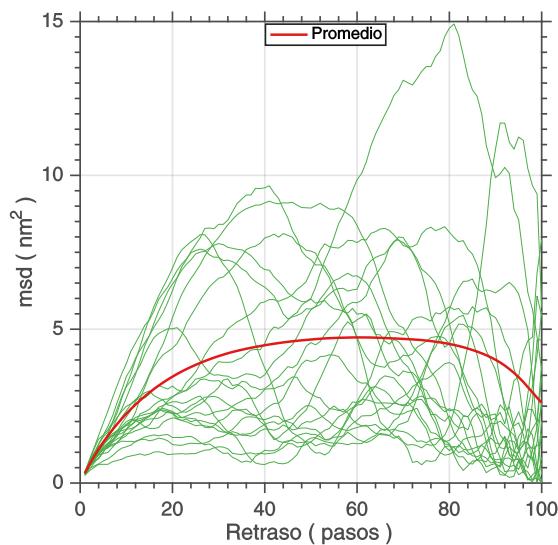


Figura 2.30: Curvas del msd para caminatas con movimiento confinado.

2. Resultados numéricos

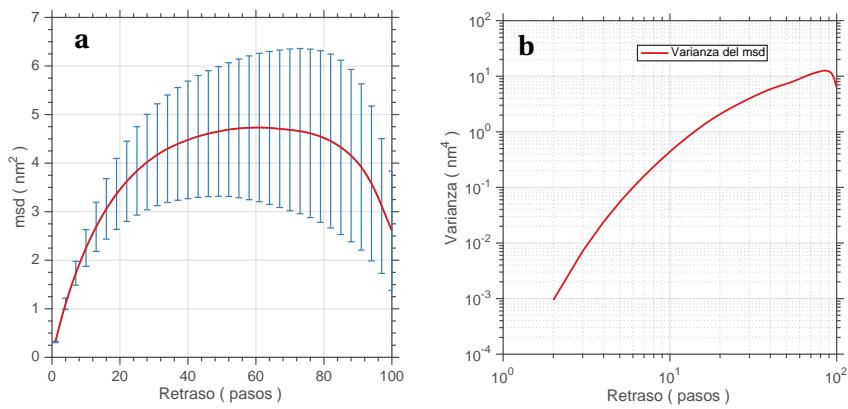


Figura 2.31: (a) Curva promedio del msd con sus barras de error de ancho iguala la desviación estándar.(b) Curva de la varianza del msd para el caso de caminatas confinadas.

3

Criterios de segmentación

En el capítulo anterior se estudió las características principales de las caminatas aleatorias cuando son generadas por un único proceso (difusión libre, confinada, con deriva y dirigida). Sin embargo en todos los casos, una inspección visual parece mostrar que los distintos tipos de movimiento están presentes a lo largo de una única caminata. Por lo tanto se ve la necesidad de generar criterios de selección binaria para distinguir caminatas libres y caminatas generadas por otro tipo de movimiento. Para luego utilizar estos criterios para segmentar trayectorias experimentales.

Los criterios de segmentación más comúnmente utilizados pueden separarse entre: los basados en propiedades locales como el cambio de dirección entre pasos y los basados en propiedades globales como la forma y el tamaño de la trayectoria. En este capítulo al criterio de segmentación lo realizamos de dos maneras. Secuencialmente de manera que el segmento crece en cada iteración del algoritmo. Y evaluando el criterio sobre una ventana que recorre toda la caminata pasando con todos los posibles tamaños de ventana. Luego se analiza el valor en función de la posición y el tamaño de la ventana.

El capítulo se divide en tres secciones. Primero se estudia la selección de caminatas dirigidas a partir del ángulo entre dos pasos consecutivos. Segundo se estudia la probabilidad de detectar caminatas confinadas a partir de la extensión espacial que tiene la caminata y su duración. Y para finalizar se estudia la forma de las caminatas, en la cual se evalúa la anisotropía de las posiciones respecto de las distintas direcciones espaciales, teniendo en mente que una caminata con deriva suele ser más alargada que una caminata libre.

Nosotros tomamos otra metodología que fue utilizar una ventana que va recorriendo sobre la trayectoria y calcula algo sobre eso y lo compara con ciertos valores de umbral.

3.1. Ángulo de giro

Para distinguir regiones en las que el movimiento aparenta ser dirigido se puede calcular punto a punto de la trayectoria el ángulo en el que la partícula cambia de dirección. Una variante de este método es el *J-analysis* introducido por Levi y Gratton en [Levi and Gratton, 2007]. En el cual, si tres pasos consecutivos el ángulo de giro se mantiene menor a un ángulo crítico se lo asigna como un punto en una región de movimiento dirigido (fig.3.1).

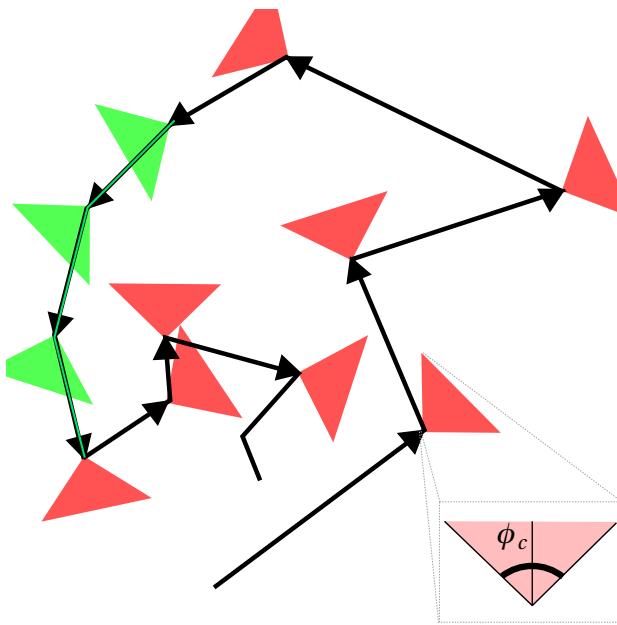


Figura 3.1: Esquema de la metodología de selección de segmentos con movimiento dirigido. En color rojo se marcan los conos en los que el ángulo entre dos pasos es mayor al ángulo crítico ϕ_c y en verde cuando es menor. Se considera como un "filamento" al conjunto de pasos consecutivos que tienen el ángulo menor al ángulo crítico.

De aquí en adelante al las regiones donde se cumple la condición de que en pasos tienen un ángulo menor al ángulo crítico las llamaremos "filamentos" (*threads*) debido a su forma alargada y en caso contrario llamaremos "segmento" a cualquier porción o pedazo de una trayectoria al cual no asignamos una característica específica.

La probabilidad de encontrar un filamento de n pasos consecutivos con un ángulo critico $0 \leq \Phi_c \leq \pi$, en una caminata con difusión libre es:

$$\begin{aligned}
 P(\phi_1 \leq \Phi_c, \phi_2 \leq \Phi_c, \dots, \phi_n \leq \Phi_c) &= \int_{-\Phi_c}^{\Phi_c} \cdots \int_{-\Phi_c}^{\Phi_c} p(\phi_1, \phi_2, \dots, \phi_n) d\phi_1 \cdots d\phi_n \\
 &= \int_{-\Phi_c}^{\Phi_c} \cdots \int_{-\Phi_c}^{\Phi_c} p(\phi_1) p(\phi_2) \cdots p(\phi_n) d\phi_1 \cdots d\phi_n \\
 &= \left[2 \int_0^{\Phi_c} p(\phi_1) d\phi_1 \right]^n \\
 &= \left[\frac{\Phi_c}{\pi} \right]^n.
 \end{aligned} \tag{3.1}$$

donde ϕ_i son los ángulos medidos entre dos pasos consecutivos y $p(\phi_i) d\phi_i$ es la probabilidad de que el ángulo ϕ_i este entre ϕ_i y $\phi_i + d\phi_i$. En el paso de la primera a la segunda línea hicimos la suposición de que los cambios de dirección en los pasos son independientes y al pasar a la tercer línea suponemos que las distribuciones son simétricas respecto de 0.

En la figura 3.2 se observa una simulación de una caminata de cinco mil pasos gaussianos junto con los filamentos (en rojo) de seis pasos, obtenidos bajo la condición de que el ángulo crítico $\Phi_c < 90^\circ$. El valor de

90° puede parecer demasiado permisivo pero lo elegimos así dado que es el límite sobre el cual se puede interpretar que dos pasos consecutivos van en el mismo sentido. A comparación del caso de caminatas en una sola dimensión donde se puede definir fácilmente el sentido del paso basándose en el signo del mismo, en dos dimensiones no es posible definir una dirección global que indique que el caminante mantiene su dirección de movimiento. Pero se puede considerar el signo de la proyección entre dos pasos consecutivos para indicar si mantiene el sentido del movimiento o lo cambió, lo que es equivalente a ver el signo del coseno del ángulo de giro. Y este cambia de signo cuando el ángulo es mayor a 90° .

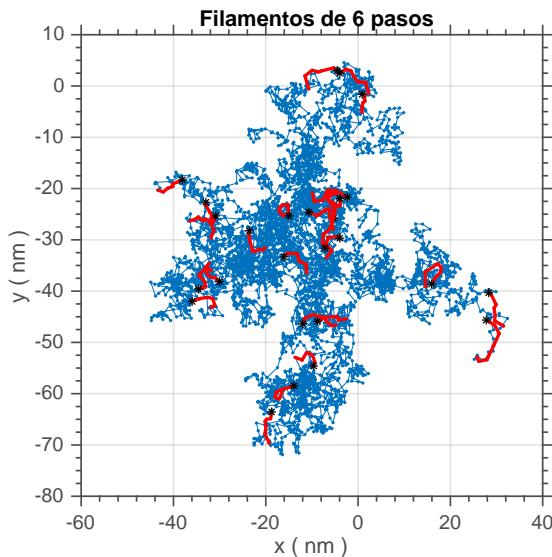


Figura 3.2: Simulación de cinco mil pasos de una caminata de pasos gaussianos $\mu = 0$, $\sigma = 1$. En color rojo se muestran todos los filamentos detectados en la trayectoria para los cuales el caminante dio giros de menos de $\Phi_c = 90^\circ$ por seis pasos consecutivos.

Para evaluar la cantidad de filamentos que se pueden encontrar en una caminata de difusión libre se simuló una caminata de cien mil pasos sobre la cual se aplicó el criterio de segmentación para detectar filamentos y se armó un histograma con el número de filamentos para cada cantidad de pasos como se muestra en la figura 3.3a. Se ajustó el histograma por una función exponencial $A \exp[-\lambda n]$ (curva roja), donde n es la cantidad de pasos del filamento y A una constante de normalización. Para ver como depende el parámetro λ con el ángulo crítico se repitió el proceso anterior para distintos valores del ángulo crítico entre 20° y 170° obteniendo que λ decrece tendiendo a cero (fig 3.3b). Para los casos donde el ángulo crítico era menor a 20° no se encontraron filamentos de más de 3 pasos por lo cual no se realizó el ajuste del histograma y el caso de 180° no tiene sentido evaluarlo.

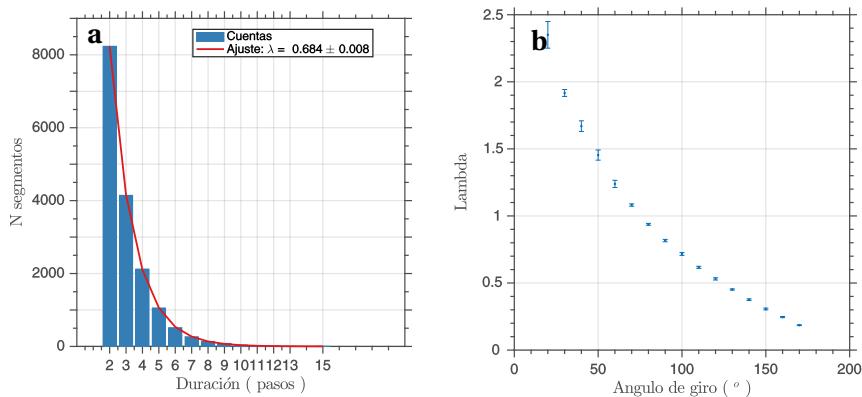


Figura 3.3: (a) Histograma del número de filamentos (giros menores a 90°) en función de la cantidad de pasos de los mismos. Los filamentos se buscaron en una caminata con cien mil pasos, donde los pasos estaban distribuidos de manera gaussiana. En rojo se grafica el ajuste del histograma por una función exponencial, el exponente ajustado fue $\lambda = 0,684 \pm 0,008$. (b) Valor del parámetro de la exponencial en función del ángulo máximo de giro impuesto para la segmentación.

Resumiendo, para buscar segmentos en los cuales hay correlación entre pasos consecutivos (los cuales llamamos filamentos), se evalúa el ángulo ϕ que existe entre la dirección de dos pasos consecutivos y si es menor a un ángulo crítico Φ_c se considera que esos pasos pertenecen a un filamento. Para construir el filamento, se evalúa este criterio paso a paso y si cumplen se lo agrega al filamento hasta encontrar un paso que no lo cumpla, donde se corta y se empieza a construir un filamento diferente.

3.2. Probabilidad de escape

A partir de la distribución de la longitud de los pasos se puede obtener la probabilidad de que la partícula se encuentre a una distancia en cierto tiempo o equivalentemente luego de cierto número de pasos. [Simson et al., 1995, Saxton, 1993b]

Esto permite determinar las regiones en dentro de la trayectoria las que la partícula difiere del movimiento browniano (de difusión libre), tanto si está confinada como si está dirigida.

Para detectar si un segmento está confinado, lo que llamaremos "cúmulo" (*cluster*), se calcula la probabilidad de que el caminante o partícula se mantenga para todo tiempo menor a t (o equivalentemente n pasos) dentro de una región de radio R alrededor del punto del cual partió durante un tiempo determinado. La condición de que el caminante se mantenga dentro de la región **para todo tiempo** (fig. 3.4b) es el detalle importante, ya que es distinto a la probabilidad de que luego de n pasos se encuentre dentro de la región (fig. 3.4a).

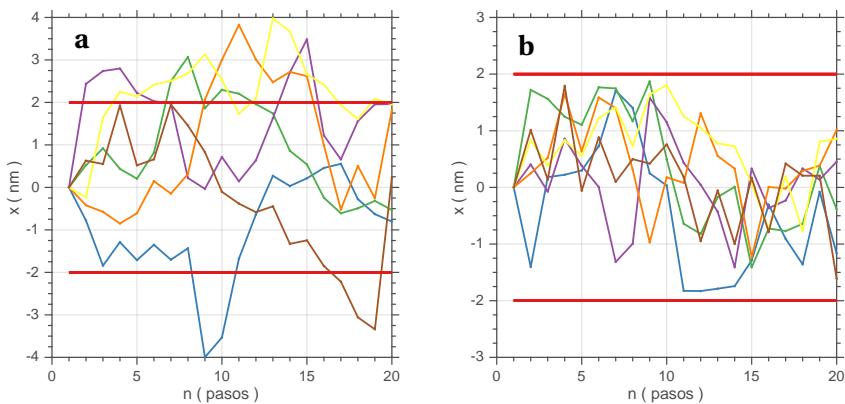


Figura 3.4: Para estudiar el confinamiento en caminatas aleatorias, estudiamos la probabilidad de que una caminata de duración n se mantenga a una distancia menor a R para **todo tiempo menor a n**

Simulando caminatas de distinta duración se calcula la proporción que cumple la condición de mantenerse en dentro de una región circular de radio R en función de la duración n de las caminatas (fig. 3.5). De manera equivalente se calcula la proporción de caminatas de una definida duración n que se mantienen dentro de la región en función del tamaño R de la misma (fig. 3.5b).

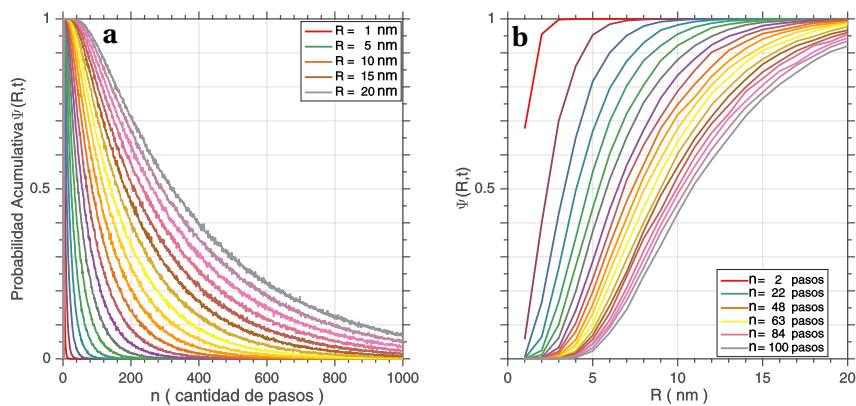


Figura 3.5: (a) Probabilidad acumulativa de encontrar una caminata que se mantenga confinada en un radio R en función del tiempo.(b) Probabilidad acumulativa de encontrar una partícula se mantenga confinada durante un tiempo t en función del radio de confinamiento R

La probabilidad acumulativa Ψ depende del parámetro $D t / R^2$, donde $D = \sigma^2 / 2\tau$ es el coeficiente de difusión, $t = \tau n$ la duración de la caminata y τ el tiempo entre mediciones. Al graficar las curvas de la figura 3.5 en función de este parámetro todas las curvas se reducen a una sola (fig. 3.6).

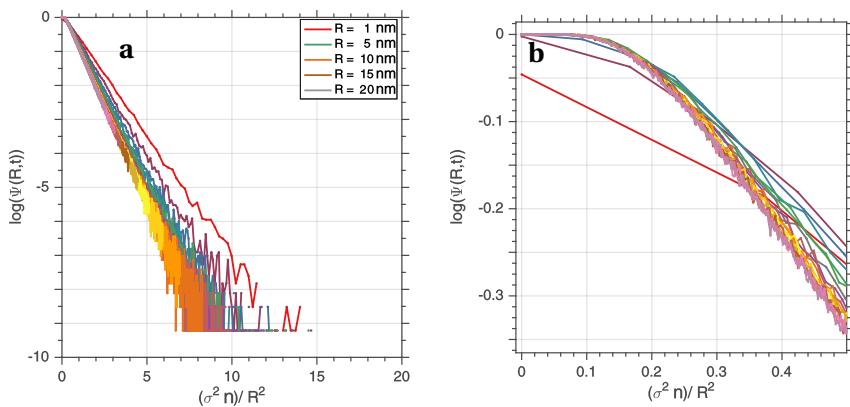


Figura 3.6: Probabilidad acumulativa en función del parámetro adimensional $\sigma^2 n/R^2$.

Se realizó un ajuste lineal sobre esta curva, considerando solo los puntos donde $\sigma^2 n/R^2 > 0,15$ obteniendo la misma curva obtenida por Saxton en [Saxton, 1993b]:

$$\log(\Psi) = (-1,163 \pm 0,004) \frac{\sigma^2 n}{R^2} + (0,2364 \pm 0,006) = (-2,325 \pm 0,009) \frac{D t}{R^2} + (0,2356 \pm 0,0063) \quad (3.2)$$

Con esto desarrollamos el algoritmo para determinar los cúmulos presentes en una trayectoria. Este consiste en seleccionar dos puntos consecutivos y evaluar el parámetro adimensional $\sigma^2 n/R^2$, utilizando σ^2 como la varianza de todos los pasos de la trayectoria, R la distancia del punto, perteneciente al segmento, más alejado al punto de partida. Definiendo una probabilidad umbral $p_c = 0,95$, se compara si $\sigma^2 n/R^2$ es mayor a $(0,236 - \log(1-p_c))/(1,163)$. Si se cumple, se agrega el siguiente punto de la trayectoria y se calcula nuevamente el radio, se evalúa el criterio. Así sucesivamente hasta que no se cumple el criterio entonces se corta el cúmulo y se empieza a considerar un nuevo cúmulo.

Con este criterio calculamos cual es la cantidad de segmentos de un dado número de pasos que se encuentran en una caminata de cien mil pasos distribuidos de manera gaussiana. Obteniendo el histograma que se muestra en la figura 3.7, el histograma se ajustó por varios modelos y el que dio el mayor coeficiente de Pearson fue con una función de la forma $A n^\alpha$, donde A es una constante de normalización y n la cantidad de pasos del cúmulo. El modelo exponencial devolvió $R_{Person} = 0,721$ mientras que la ley de potencia $R_{Person} = 0,998$.

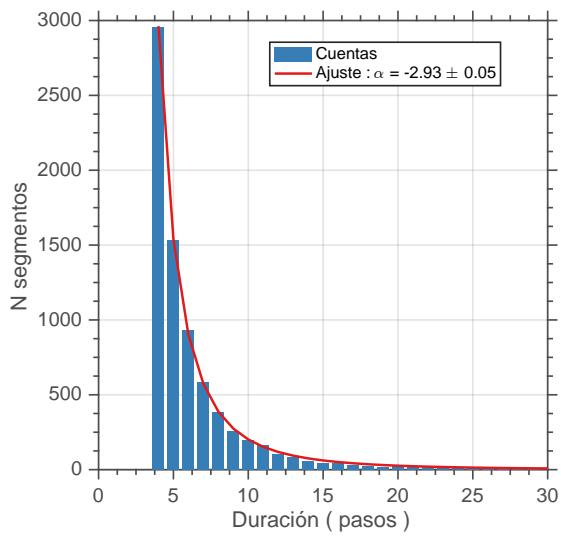


Figura 3.7: Histograma del número de cúmulos encontrados en una caminata de cien mil pasos. El ajuste se realizó por una función de la forma $A n^\alpha$, donde A es una constante de normalización y n la cantidad de pasos del cúmulo.

En la figura 3.8 mostramos los cúmulos en la misma caminata de la figura 3.2 en la cual se marcaron los filamentos.

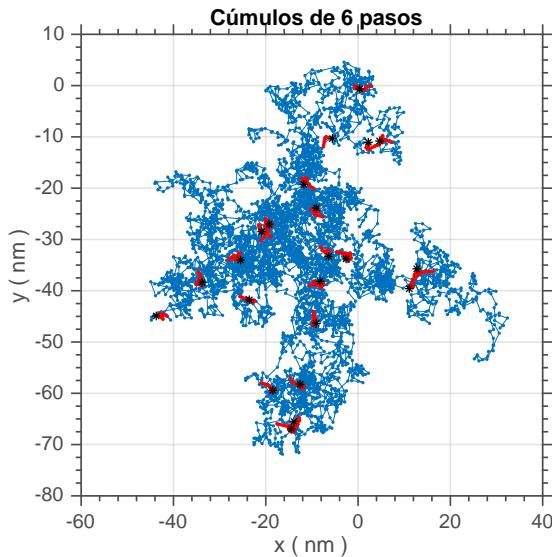


Figura 3.8: Simulación de cinco mil pasos de una caminata de pasos gaussianos. La caminata es la misma presentada anteriormente en la figura 3.2. En rojo se marcan los cúmulos obtenidos a partir del cálculo de la probabilidad de escape.

3.3. Forma de la caminata

Una característica simple de evaluar en las caminatas y que a menudo se utiliza para sacar conclusiones globales de la misma es la "forma" de la caminata. Una manera de definir cuantitativamente la "forma" de una caminata es a través del tensor de giro. En un trabajo de Saxton [Saxton, 1993b] se muestra cómo utilizar este parámetro para distinguir caminatas con deriva. Es un complemento de la probabilidad de escape estudiada en la sección 3.2 que distinguía a partir de la función de probabilidad las caminatas confinadas en una región. Se puede observar en las simulaciones que las caminatas con deriva tienden a ser más alargadas y las caminatas confinadas son más compactas.

El tensor de giro definido como ec.3.3[Rudnick and Gaspari, 1986] describe el tamaño de la caminata en cada dirección y la relación que existe entre las distintas direcciones de movimiento.

$$G_{ij} = \frac{1}{N} \sum_{k=1}^N (x_{ik} - \langle x_i \rangle)(x_{jk} - \langle x_j \rangle), \quad (3.3)$$

En la ec.3.3 el promedio se realiza sobre todas las posiciones en las que estuvo el caminante. Reescribiendo de manera matricial para el caso bidimensional:

$$G = \begin{bmatrix} \langle x^2 \rangle - \langle x \rangle^2 & \langle xy \rangle - \langle x \rangle \langle y \rangle \\ \langle xy \rangle - \langle x \rangle \langle y \rangle & \langle y^2 \rangle - \langle y \rangle^2 \end{bmatrix} \quad (3.4)$$

El tensor es equivalente a la matriz de covariancia y describe la extensión promedio en cada una de las direcciones (términos de la diagonal) y la relación lineal que existe entre las distintas direcciones.

Al ser simétrico y definido positivo se puede diagonalizar y sus autovalores son reales y positivos. Los autovectores representan dos direcciones ortogonales donde la extensión de la caminata es mínima y máxima. Y la raíz cuadrada de los autovalores son las distancias promedio en estas direcciones. En la base que diagonaliza la matriz se tiene:

$$\hat{G} = \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix} = \begin{bmatrix} R_1^2 & 0 \\ 0 & R_2^2 \end{bmatrix} \quad (3.5)$$

Consideramos los autovalores de menor a mayor de forma que $\lambda_2 \leq \lambda_1$.

La información que tiene el tensor esta en los autovalores, uno puede formar distintos parámetros en función de los autovalores para describir tanto la forma y el tamaño de la caminata.

La extensión de la caminata se mide a través del radio de giro que está definido por:

$$R_g^2 = R_1^2 + R_2^2 = \lambda_1 + \lambda_2 = \text{Tr}(G) \quad (3.6)$$

Se conoce una relación entre el $\text{msd}(n)$ y el radio de giro cuando las caminatas tienen una longitud de paso fija[Gordon, 1972]:

$$\langle R_g^2(n) \rangle = \frac{1}{6} \left[1 - \frac{1}{n^2} \right] \text{msd}(n). \quad (3.7)$$

En el límite de muchos pasos, este tipo de caminata se asemeja a caminatas libres donde la distribución se aproxima a una gaussiana, de manera que en este límite el radio de giro es proporcional al $\text{msd}()$.

En la figura 3.9 mostramos una caminata junto con la elipse definida por el tensor de giro y un círculo de radio R_g . Hay que notar que el radio de giro siempre es mayor a los radios principales debido al Teorema de Pitágoras, de manera que la elipse siempre está contenida dentro del círculo de radio R_g .

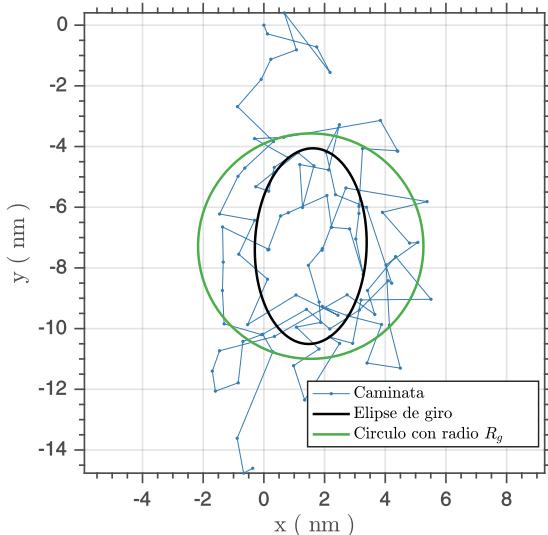


Figura 3.9: Caminata junto con su elipse de giro y círculo de radio R_g . Así definidas, la elipse siempre está contenida en el círculo de radio R_g .

Otra manera muy utilizada de combinar los autovalores del tensor de giro y que permite concluir sobre la forma global de la caminata es la anisotropía, definida como el cociente entre el autovalor más pequeño y el más grande.

$$a = \frac{\lambda_2}{\lambda_1}, \quad (3.8)$$

La anisotropía toma valores entre 0 y 1 lo cual nos independiza de la escala del sistema. En la figura 3.10 se muestran caminatas con distintos valores de anisotropía junto con la elipse que se obtiene a partir del tensor de giro.

3. Criterios de segmentación

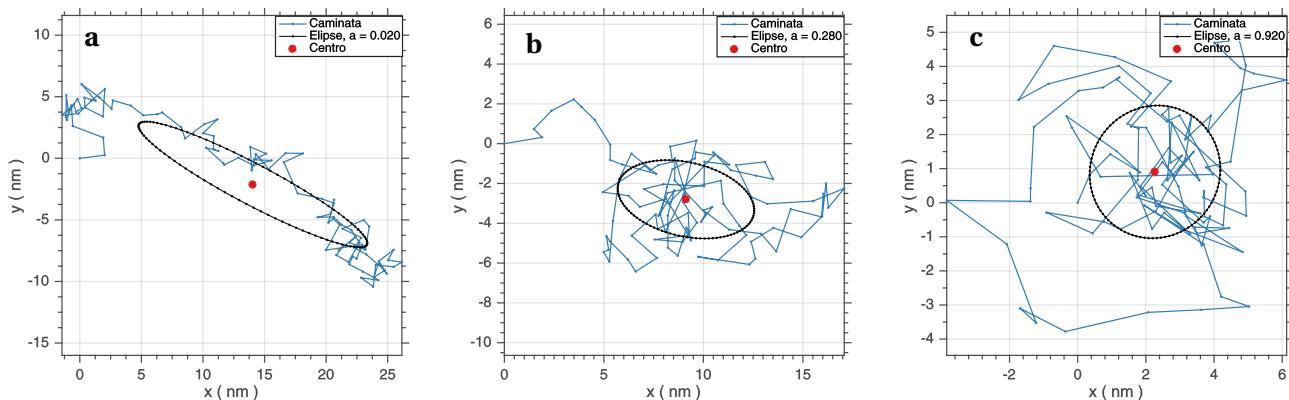


Figura 3.10: Ejemplos de caminatas con distintos valores de anisotropía. (a) Caminata con anisotropía menor a 0.03. (b) Caminata con anisotropía próxima al valor medio de la anisotropía para caminatas libres 0.28. (c) Caminata con anisotropía mayor a 0.8.

La distribución de anisotropía es independiente del número de pasos de la caminata, para $n > 10$ (3.11b). Esto nos permite generar un criterio de selección independiente del mismo. Definiendo un valor umbral de anisotropía podemos separar los casos en los que las caminatas son alargadas debido a la presencia de una deriva o que el movimiento sea dirigido.

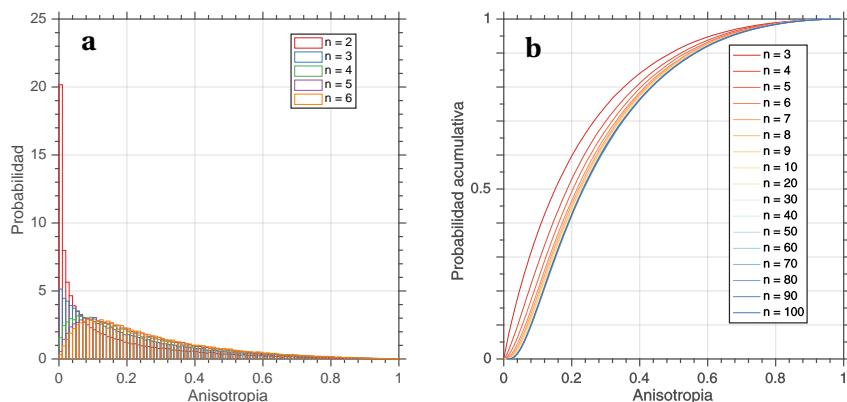


Figura 3.11: (a) Histogramas de la probabilidad del valor de anisotropía para caminatas Normales para distintas duraciones de la caminata. (b) ECDF(*Función empírica de probabilidad acumulativa*) de la anisotropía para distintas duraciones de la caminata, se puede observar como las curvas convergen en pocos pasos

La forma funcional de la distribución de probabilidad de la anisotropia fue calculada por Sciutto en [Sciutto, 1995, Sciutto, 1994], para los casos de caminatas con longitud de pasos fija y bajo la condición de tener una gran cantidad de pasos $n > 100$. Bajo la condición de tener un gran numero de pasos, la distribución de probabilidad de encontrar al caminante a una distancia definida converge al caso browniano por lo cual se puede utilizar la misma función en casos de caminatas con pasos gaussianos de menor duración. La distribución de probabilidad de la anisotropia esta definida por:

$$P_a(z) = \frac{1}{N} \frac{z^{w-1}(1-z)^{\gamma-1}}{(1+\beta z)^{w+\nu}} \quad 0 \leq z \leq 1 \quad (3.9)$$

Los parámetros no tienen ninguna interpretación clara ya que provienen de combinación de parámetros de las distribuciones de probabilidad de los autovalores. Para ajustar la distribución al modelo propuesto definimos una función alternativa que consiste en la integral hasta el valor especificado. Y entregándola a funciones del modulo de optimización de Matlab obtenemos que los valores de ajuste son los siguientes. Para independizarnos del tamaño del bineado y no ajustar a un histograma de la distribución ajustamos la integral a la ecdf. Y con los valores ajustados graficamos la distribución junto con un histograma de los datos.

Realizando un ajuste sobre la probabilidad acumulativa obtenemos los parámetros:

$$\beta = 29,50, \quad \gamma = 2,09, \quad \nu = 0,32, \quad w = 6,45. \quad (3.10)$$

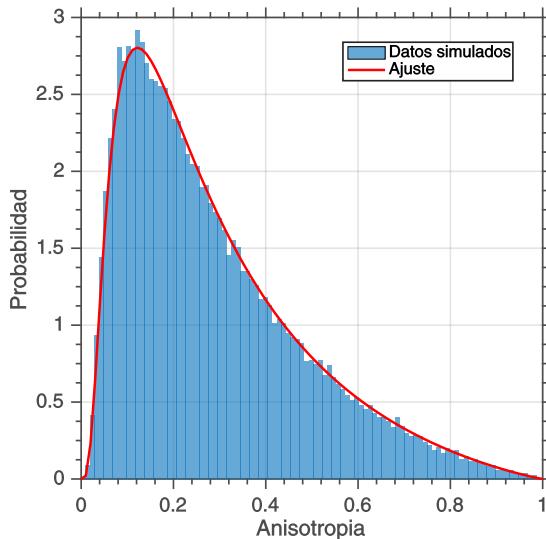


Figura 3.12: Histograma de la anisotropía calculada a partir de caminatas simuladas de $n = 100$ pasos y pasos distribuidos de manera gaussiana. En rojo la distribución de probabilidad dada por la ec. 3.9 utilizando los parámetros ajustados.

Algo que no se tiene en cuenta al evaluar la forma de toda la caminata a partir del histograma de la distribución de los pasos es el orden temporal en el que ocurren los pasos. Por lo cual implementamos un algoritmo de selección en función de la anisotropía sobre ventanas de distinto tamaño que recorren toda la caminata. Como se puede ver claramente en la figura 3.13 la distribución de anisotropía para caminatas que tienen deriva es más angosta y tiende a estar más pegada hacia los valores cercanos a cero. Esta tendencia se acentúa medida que las caminatas tienen más pasos. Este resultado sugiere que la anisotropía nos indica que es un buen parámetro para detectar movimiento dirigido.

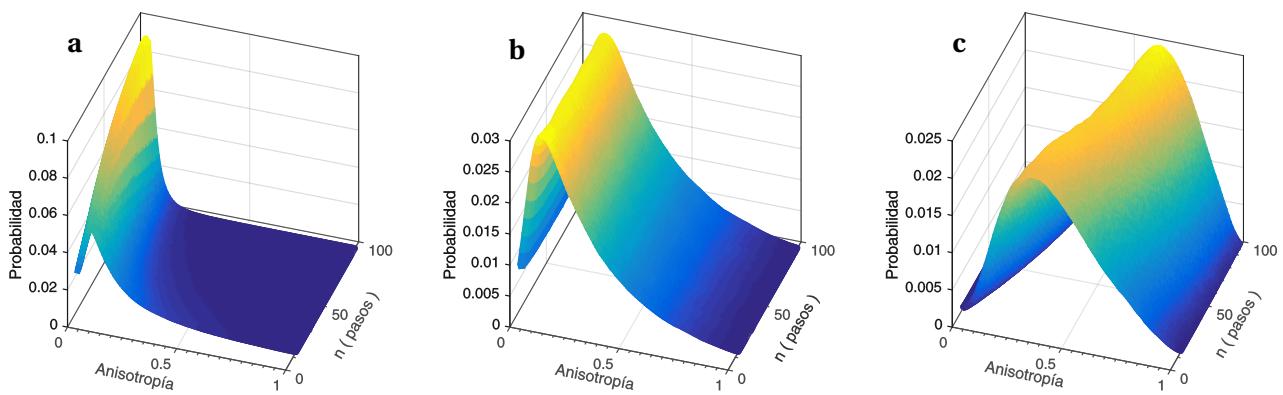


Figura 3.13: Superficies de la distribución de la anisotropía en función del número de pasos de la caminata para: (a) Caminatas con deriva, (b) caminatas libres y (c) para caminatas confinadas.

3.4. Criterio de selección

La figura 3.13 demuestra que el valor de anisotropía depende fuertemente del tipo de movimiento y es independiente de la cantidad de pasos cuando el movimiento es libre. Por este motivo es necesario postular un valor umbral de anisotropía a_0 de manera que toda caminata que tenga anisotropía menor al umbral es considerada como caminata con deriva.

Es necesario cuantificar las tasas de falsos positivos y verdaderos positivos para poder evaluar la precisión del criterio de selección. En la figura 3.14a las curvas representan la fracción de caminatas que son consideradas como caminatas con deriva utilizando un valor umbral de $a_0 = 0,04$ que corresponde al 1-percentil de la distribución de anisotropía. Hay cuatro parámetros de los cuales depende el método: la anisotropía umbral, la velocidad de deriva o valor medio de los pasos, la cantidad de pasos de la caminata o duración y el coeficiente de difusión. Los tres últimos se pueden agrupar en un único parámetro adimensional como $\gamma = v^2 t / D$. Luego, expresando las probabilidades en función γ todas las curvas se reducen a una sola (fig. 3.15b).

Para estudiar el criterio en función de la anisotropía umbral graficamos las curvas ROC (fig. 3.14b) donde se comparan la tasa de verdaderos positivos contra la tasa de falsos positivos. En esta curva cada punto indica los resultados obtenidos para un valor de la anisotropía umbral. El valor que uno querría tener en el criterio es el de $(0, 1)$ que indica el 100% de detección sin falsos positivos. Se puede seleccionar el punto de la curva más cercano a $(0, 1)$ como el punto que da los mejores resultados (puntos negros) y con esto se selecciona el valor de anisotropía umbral óptimo. Con esto uno puede definir la probabilidad de estar equivocado al seleccionar una caminata como una con deriva.

Para segmentar la caminata utilizamos este criterio mientras barremos con una ventana de tamaño $L_{ventana}$, un número impar. Graficamos en un mapa de temperatura los valores de anisotropía obtenidos en función del tamaño de la ventana y la posición del centro para observar las regiones que se pueden considerar como segmentos con deriva.

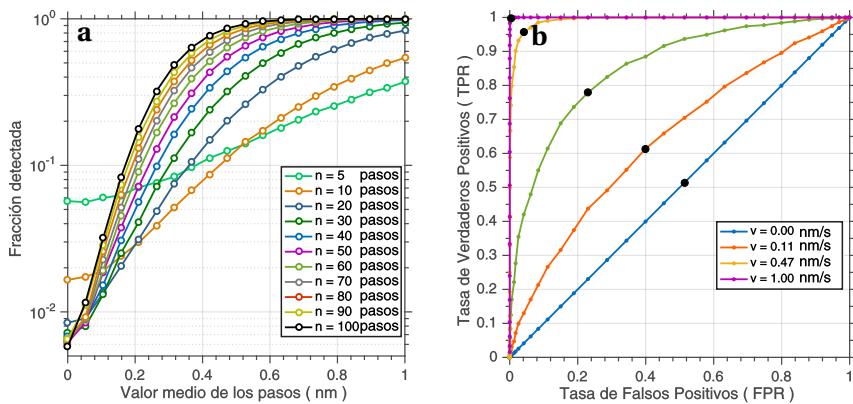


Figura 3.14: (a) Curvas de la fracción de caminatas detectadas como caminatas con deriva en función del valor medio de los pasos, para distintas duraciones de las caminatas. (b) Curvas ROC del criterio de selección de caminatas con deriva, las curvas se generaron variando el valor crítico de la anisotropía. El punto negro indica el valor más cercano a (0,1) e indica el punto de mejor comportamiento del algoritmo de selección.

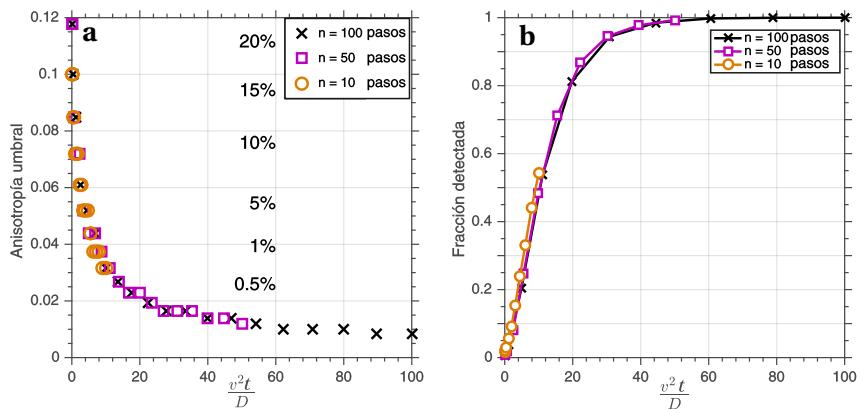


Figura 3.15: (a) Valores del umbral de la anisotropía sobre los cuales el criterio de selección tiene el mejor comportamiento. Estos valores corresponden a los puntos negros de la figura 3.14b . (b) Tasa de detección de caminatas con deriva en función del parámetro adimensional $\gamma = v^2 t / D$

3.4.1. Metodología de análisis

Para segmentar las caminatas definimos el *tensor de giro local* como:

$$\hat{\mathbf{g}}_{ij} = \frac{1}{n_2 - n_1} \sum_{l=n_1}^{n_2} (x_{il} - \langle x_i \rangle)(x_{jl} - \langle x_j \rangle), \quad (3.11)$$

en la cual solo se consideran los puntos de la caminata entre los pasos \$n_1\$ y \$n_2\$.

Para representar todos los valores de anisotropía local que se obtienen en una única caminata se generando un mapa en la que se indica con distintos colores el valor de la anisotropía en función de la posición del segmento que se evalúa y la cantidad de pasos del mismo (fig. 3.16).

3. Criterios de segmentación

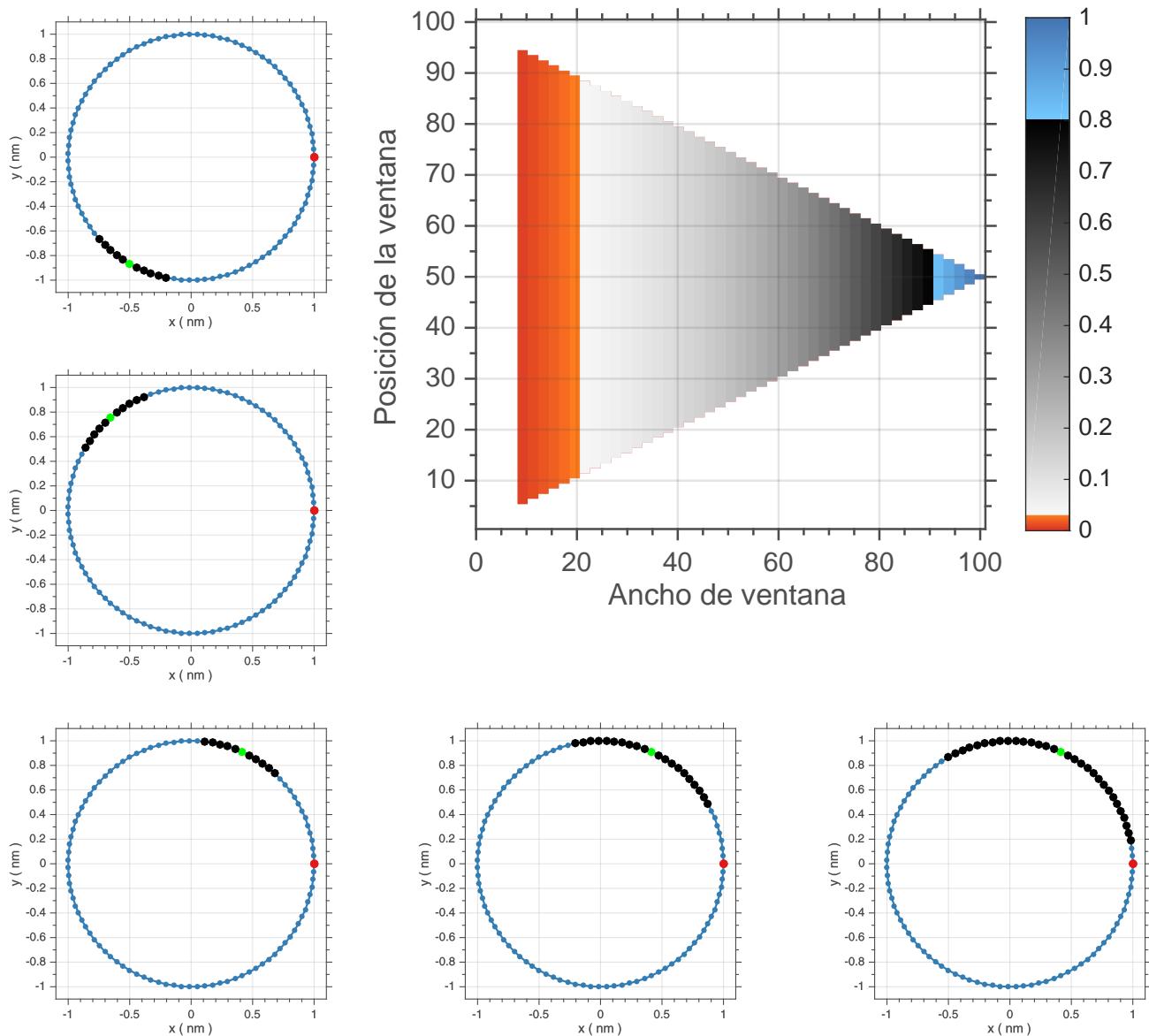


Figura 3.16: Figura explicativa del mapa de colores donde se representa la anisotropía local. En la columna de la izquierda se muestra como la ventana va recorriendo distintas posiciones en la caminata y en la fila inferior se muestra como va aumentando el tamaño de la ventana. Junto con el mapa de colores para una caminata que consiste de 100 puntos ubicados sobre un círculo.

Cuando la forma de la trayectoria es más compleja en el mapa se observan regiones semejantes a islas en tonos rojos que representan las regiones de deriva.

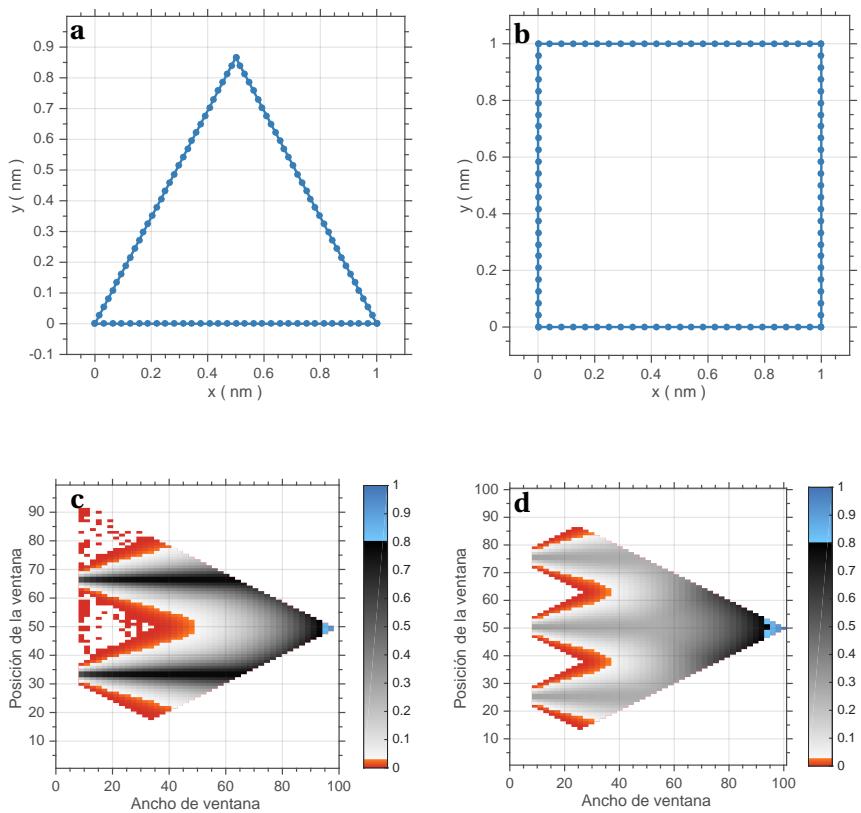


Figura 3.17: Caminatas de 100 puntos repartidos en figuras geométricas simples. Debajo de éstas, los mapas de anisotropía correspondiente a las mismas

Para interpretar los mapas se siguen las siguientes reglas. Nos vamos a referir a las regiones en las cuales hay una clara división de colores con su entorno como islas. islas son regiones de un color acotadas tal que su entorno es de otro color. El vértice superior(inferior) de una ventana como el que tiene un índice mayor(menor) al centro de la ventana

En las islas en las que crece el tamaño de la ventana y retrocede la posición de la ventana, el vértice superior está fijo. En las islas en las que crece el tamaño de la ventana y avanza la posición de la ventana, el vértice inferior está fijo. Las islas las cuales se angostan a medida que crece el tamaño de la ventana indican regiones en las que la ventana más grande incluye a todas las ventanas más pequeñas. Es decir, la punta indica el segmento más grande. Las islas en las que no aumenta la posición de la ventana indican el centro del segmento.

Como ejemplo se puede observar en los mapas del triángulo y el cuadrado los bordes superiores rojos/anaranjados en las "V" indican regiones con anisotropía cercana a cero es decir figuras alargadas. Estas islas empiezan cerca de los múltiplos de 30 que son las posiciones de los vértices y tienen su extremo en la punta que está en la posición media entre dos vértices. Entonces si el vértice superior de la ventana está fijo, a medida que la ventana crece la ventana tiene que retroceder. Para el segundo caso las islas inferiores en las "V" crecen marcando el crecimiento de la ventana pero fijando el extremo inferior. Las islas negras horizontales en el mapa del triángulo marcan ventanas centradas en los vértices del triángulo y crecen hasta un tamaño cerca de 60 indicando que cubren completamente las dos aristas que nacen de ese vértice y su anisotropía está es casi constante ya que la figura que tienen esos segmentos es la misma, y lo único que cambia es el tamaño, para lo cual la anisotropía es indistinta

3. Criterios de segmentación

Hay un problema en estos ejemplos y es que las aristas son líneas rectas por lo tanto tienen anisotropía nula y en el mapa no se grafican estos puntos, en el por lo que los triángulos blancos son regiones de anisotropía nula. En el caso de caminatas aleatorias se sabe que no pueden tener anisotropía nula por lo tanto no aparece este problema. La elección de graficar los mapas con los tres gradientes es por completitud del mapa y es más fácil seguir las tendencias de las islas si se incluye la región de gradiente blanco-negro.

Mapa de caminatas simuladas

Para finalizar este capítulo mostramos una caminata de mil pasos junto con un segmento obtenido mediante el criterio de la anisotropía. A la par se muestra el mapa correspondiente a la caminata donde se puede observar como los segmentos con deriva son muy poco frecuentes y con duraciones del orden de las decenas de pasos.

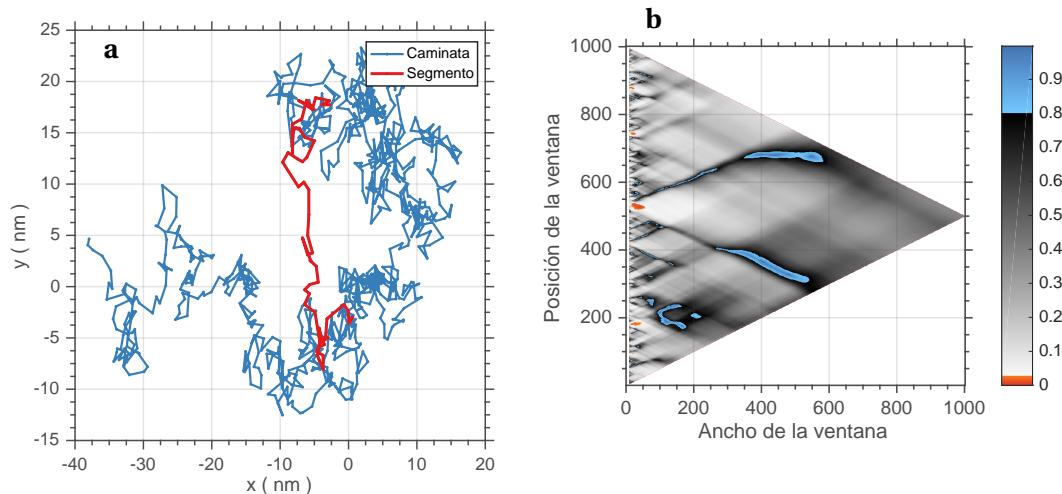


Figura 3.18: (a) Mapa para una caminata gaussiana simulada, junto con el segmento correspondiente a la isla naranja cercana a la posición 500 dentro de la caminata.(b) Mapa de anisotropía correspondiente a la caminata.

4

Análisis de trayectorias experimentales

*“We usually see only the things we are looking for,
so much that we sometimes see them where they are not”*
Eric Hoffer

Presentamos una metodología para analizar caminatas individuales basada en los resultados obtenidos en la sección de Caminatas Aleatorias 2.1, donde vimos cómo los distintos tipos de movimiento se caracterizaban por los momentos de la distribución de pasos y de la relación que existía entre los pasos.

Analizando la distribución de los pasos vamos evaluar mediante test estadísticos las distintas características del movimiento y damos una guía para interpretar que puede ocasionar desviaciones del movimiento esperado. El análisis lo realizamos para el caso de caminatas bidimensionales pero es posible extender los resultados a otras dimensiones.

Cómo todo análisis vamos a presentar varias formas de representar los datos que permiten al usuario reconocer características fundamentales en el movimiento y a la par de esto, damos las herramientas para el análisis cuantitativo.

La motivación de este trabajo consistió en analizar cuál es la información que se puede obtener de caminatas aleatorias individuales dado que en el Laboratorio de Electrónica Cuántica se realizan experimentos de seguimiento de partícula única (SPT) aplicados a biología celular.

4.1. Dispositivo experimental

Los métodos de seguimiento de partículas individuales (*single particle tracking techniques*, SPT) permiten determinar la posición de una única partícula en función del tiempo con una resolución espacial que depende básicamente del número de fotones colectados. El procedimiento de SPT más comúnmente utilizado consiste en tomar una secuencia de imágenes haciendo un barrido lineal con el láser de excitación mientras la partícula se mueve. Luego se determina la posición de la partícula en cada imagen y así se reconstruye la trayectoria. Si bien este procedimiento puede resultar útil para estudiar la dinámica de muchos sistemas, al tener que tomar una imagen completa en cada paso, la velocidad de todo el proceso es un limitante. Otra de las limitaciones importantes de las técnicas basadas en la adquisición secuencial de imágenes es que permiten estudiar sólo el movimiento en el plano de la imagen. Una técnica alternativa que apareció hace algunos años es el seguimiento de partículas por barrido orbital. Desarrollada por Kis-Petrićova y Gratton en 2004 [Kis-Petrićova2004] fue originalmente aplicada al monitoreo de la dinámica de microesferas fluorescentes en solución, y rápidamente extendida al estudio de procesos dinámicos dentro de una célula viva [Levi2005], [Hell2008]. Sin entrar en detalles técnicos, durante el barrido orbital se mide la intensidad de fluorescencia en 2^n puntos a lo largo de la órbita circular de la cual se puede extraer información sobre la distancia de la

superficie del objeto al centro de la órbita y la posición angular del objeto respecto de la órbita. En la figura 4.1 se muestra esquemáticamente el principio de funcionamiento del método.

Cuando el objeto que se desea seguir está centrado respecto del centro de la órbita generada por el barrido con el haz de excitación, la intensidad medida a lo largo de la órbita debe ser constante. Si, en cambio, el objeto no está centrado respecto de la órbita, la intensidad detectada estará modulada por la distancia del objeto a la órbita, alcanzando un máximo de intensidad en la posición en la que la partícula intersecta la órbita. En este trabajo analizamos trayectorias experimentales obtenidas a partir del microscopio multifotónico por barrido orbital construido íntegramente en el LEC [Gabriel, 2015].

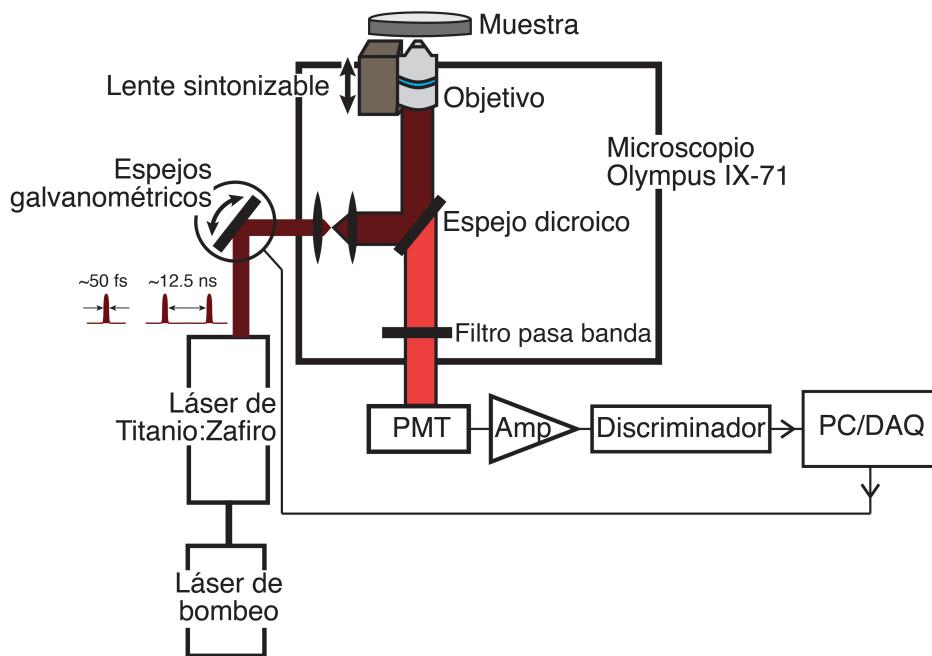


Figura 4.1: Esquema del principio de funcionamiento del método de seguimiento de partículas por barrido orbital. Figura extraída de [Gabriel, 2015]

4.2. Trayectorias experimentales

En la figura 4.2 se muestran trayectorias medidas de nanopartículas individuales de oro de 10 nm de diámetro (AuNP) difundiendo dentro del núcleo de células de ovario de hámster chino (CHO-K1). Estas mediciones fueron realizadas por la Lic. Manuela Gabriel en el instrumento y con el método recién descrito. No fue parte de este trabajo de tesis realizar las mediciones experimentales y sí fue trabajo su análisis e interpretación a partir de los modelos propuestos.

En ellas se observa cómo la partícula realiza distintos tipos de movimiento. Hay regiones en las que parece que la partícula pasa más tiempo y regiones en las que el movimiento parece estar dirigido.

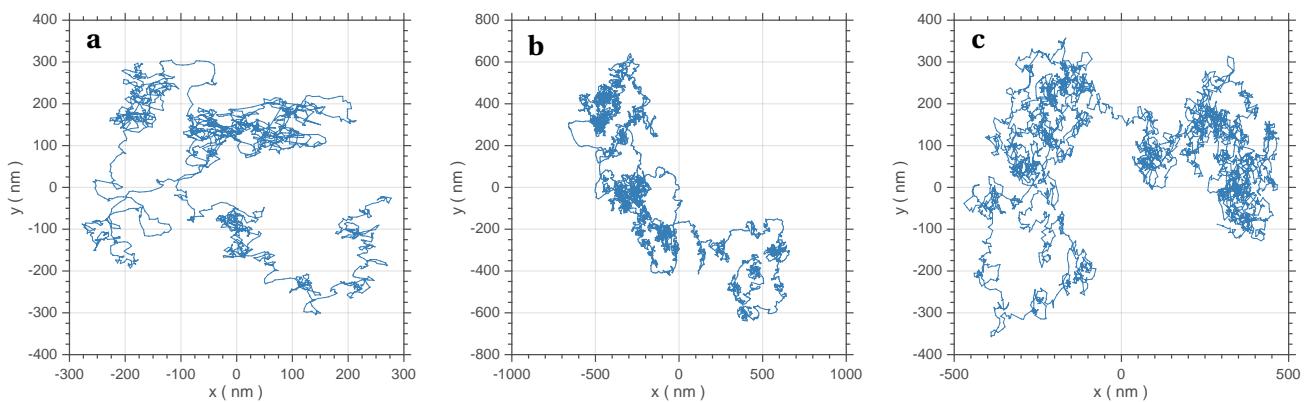


Figura 4.2: Trayectorias de nanopartículas de oro difundiendo en el interior del núcleo de células de ovario de hámster chino (CHO-K1). La duración de los experimentos fue de a) 45 min, b) 38 min y c) 50 min.

Hasta el momento en este trabajo analizamos los casos en donde la partícula un único tipo de movimiento durante toda la trayectoria. Sin embargo, como se desprende de la figura 4.2 en las caminatas experimentales esto no ocurre, o al menos parece no ocurrir. Es central entonces tener un método que permita diferenciar distintos tipos de movilidad de manera automatizada en distintas trayectorias o incluso determinar las partes dentro de una misma trayectoria donde ocurre un cambio en el tipo de movimiento. A partir de los resultados obtenidos en los capítulos anteriores trabajando con caminatas simuladas numéricamente en la sección siguiente proponemos un protocolo para analizar trayectorias experimentales y cómo identificar si realmente existen distintos movimientos o solo son artefactos debidos a la aleatoriedad del movimiento. Además estudiamos distintos algoritmos de segmentación que permiten detectar regiones donde el movimiento no se comporta como movimiento aleatorio libre.

4.3. Análisis de trayectorias experimentales

A modo de resumen aplicamos los resultados obtenidos hasta el momento a un problema concreto como puede ser la determinación de los puntos dentro de una trayectoria experimental donde hay un cambio en el tipo de movilidad de la partícula. Proponemos un protocolo detallado para el análisis de las trayectorias experimentales. El análisis aquí propuesto puede hacerse con las rutinas escritas para este fin en entorno Matlab en el marco del trabajo de esta tesis, y códigos que pueden obtenerse en el repositorio `randomWalk-matlab` en Github[Kielbowicz, 2017].

El análisis de las trayectorias consiste básicamente en realizar un análisis estadístico de los pasos. Considerando los pasos como una variable aleatoria, hay que determinar la distribución a la que pertenecen los pasos, si el movimiento es independiente en cada dirección, si existe una dirección preferencial para dar los pasos o si el movimiento es homogéneo y si hay una relación entre pasos consecutivos.

Lo primero a determinar es la distribución de la que provienen los pasos. Se espera que la distribución sea Normal si todos los pasos fueron generados de manera independiente y provienen de la interacción de la partícula con el medio. Las diferentes desviaciones de ésta distribución nos van a indicar distintas cosas del movimiento. Para analizar la distribución de la que provienen los pasos es necesario determinar si el movimiento se puede descomponer en las distintas dimensiones ya que a partir de eso uno define bien la variable aleatoria que va a analizar.

En el caso bidimensional uno tiene que la variable aleatoria es un vector de dos componentes $X = [X, Y]$

donde cada componente es una variable aleatoria. Para evaluar la distribución uno puede realizar un gráfico de dispersión donde se grafican todos los pasos partiendo desde el origen como se muestra en el gráfico polar en la figura 4.3a. A partir de este gráfico uno puede detectar si existe una correlación entre el movimiento en distintas direcciones. Algunos de los casos con los que uno se puede encontrar son los siguientes: Si no hay correlación lo que uno observa es una nube de puntos con simetría circular, lo que indica que no existe una dirección preferencial del espacio(movimiento isótropo). Si se observa una nube de puntos pero con perfil elíptico el movimiento es anisótropo, es decir, no es lo mismo moverse en una dirección u en otra. Esta elipse puede estar alineada con los ejes de medición o tener cierta inclinación, que se puede evaluar con la correlación. Esto no es un inconveniente ya que se puede realizar una rotación para alinear los ejes con los ejes principales de la elipse. Otro caso puede ser que la nube tiene forma de cruz, lo que indicaría que el movimiento se está realizando en una red dado que es posible moverse en una única dirección en cada paso.

La diferencia principal entre los dos primeros casos y el ultimo es que es más probable que para los primeros, la distribución conjunta se pueda descomponer como el producto de las distribuciones marginales, indicando que el movimiento es independiente en cada dirección. Esto es posible en un único caso, si el movimiento en cada dirección es gaussiano.

Recordando que transformaciones lineales de vectores aleatorios gaussianos, siguen siendo gaussianos y por el siguiente teorema[Hogg et al., 2014]:

Teorema. *Sean dos variables aleatorias X, Y . Si siguen una distribución gaussiana bivariada $N(\mu, \Sigma)$, entonces:*

Son independientes si y sólo si la correlación es nula.

únicamente en el caso de que en ambas direcciones la distribución sea gaussiana, implica que los pasos en cada dirección son independientes.

En esta tesis sólo se presentan herramientas para el caso en el que se descompone el movimiento. En este caso hay una gran variedad de herramientas estadísticos para evaluar sobre cada componente de los pasos. Y en el caso de que no se puede descomponer, solo se evaluaron los valores medios.

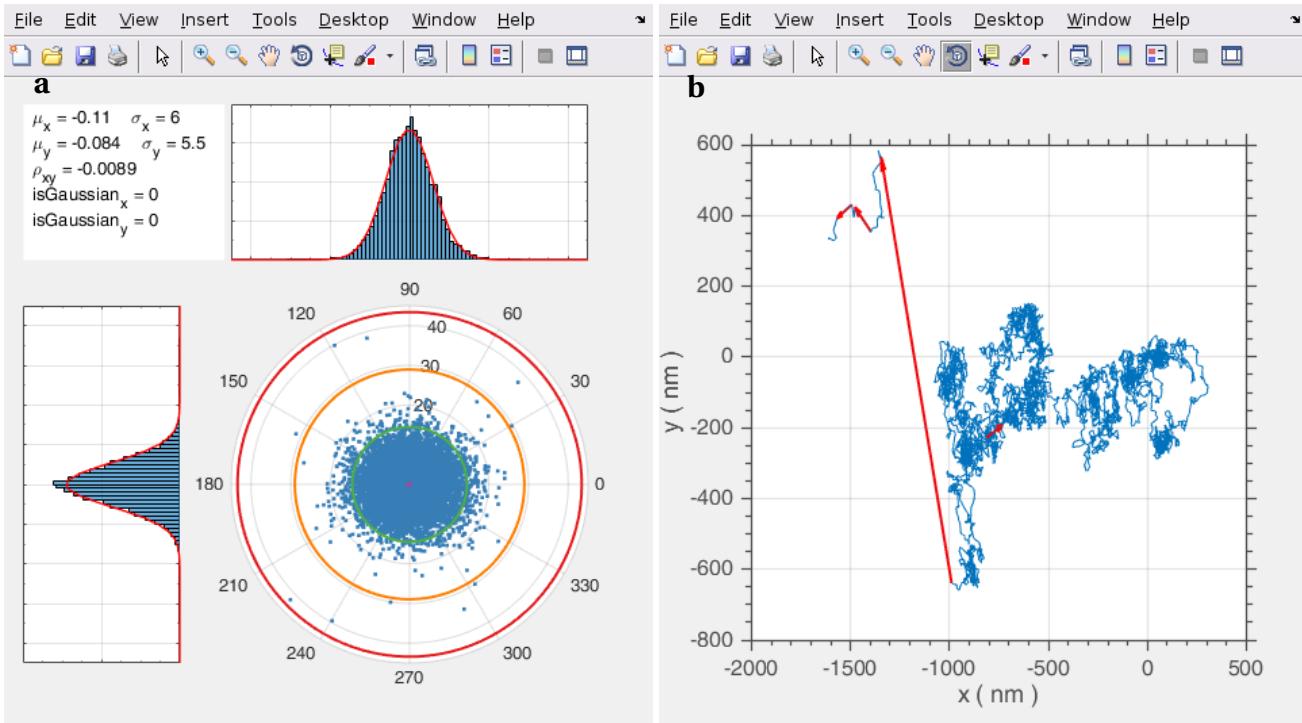


Figura 4.3: Gráficos realizados por la función `plotScatter()`. **a)** Gráfico polar de los pasos que componen la trayectoria, se marcan los puntos correspondientes a una desviación estándar $\sigma = \sqrt{\sigma_x^2 + \sigma_y^2}$ (verde), 2σ (naranja) y 3σ (rojo). A la izquierda (resp. arriba) del gráfico polar, se grafica el histograma de la proyección de los pasos sobre el eje y (x) junto con la distribución Normal, con parámetros estimados de los datos. En el recuadro se indican los valores de los parámetros estimados donde μ es el valor medio y ρ_{xy} la correlación. Las variables `isGaussian` indican si se acepta la hipótesis nula de que los datos provienen de una distribución normal, evaluado con un test de Lilliefors. **b)** Gráfico de los trayectoria experimental junto con los pasos (rojo) de longitud mayor a 3σ los cuales son considerados como *outliers* y se descartan en el análisis.

Al analizar los pasos uno busca extraer información cuantitativa del movimiento de la partícula. Como se muestra en el capítulo 2 tanto el valor medio y la varianza de los pasos proveen la información del movimiento. La varianza σ^2 está relacionada con el coeficiente de difusión D que indica el nivel de actividad del medio y el valor medio μ ésta relacionado con la velocidad de deriva v que indica que hay una dirección de preferencia para el movimiento. Dado el tiempo τ entre mediciones de la posición, las relaciones son las siguientes

$$D = \frac{\sigma^2}{2\tau}, \quad v = \frac{\mu}{\tau}$$

Los valores obtenidos de esta manera son estimaciones puntuales de los parámetros reales y no proporcionan un intervalo de confianza sobre los mismos. Para calcular los intervalos de confianza se pueden utilizar las fórmulas:

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}, \quad \sigma_S = \sqrt{\frac{\mu_4 - \sigma^4}{4n\sigma^2}}, \quad \sigma_{S^2} = \sqrt{\frac{\mu_4 - \sigma^4}{n}} \quad (4.1)$$

Hacemos la aclaración de que en estas fórmulas la notación se vuelve repetitiva y puede volverse un poco complicado entender qué representan exactamente. σ^2 y μ_4 son la varianza y el cuarto momento central de la

distribución de la cual se estiman los parámetros \bar{X} y S que son el valor medio y la varianza muestral, es decir se calculan a partir de una muestra de n elementos. Y tanto $\sigma_{\bar{X}}$, σ_S y σ_{S^2} son las desviaciones estándar de los estimadores muestrales, las cuales definen los intervalos de confianza (para resolver dudas ver el capítulo de estadística en [Spiegel et al., 2012]). Lo característico de estas desviaciones es la dependencia con la cantidad de elementos en la muestra, haciendo que a mayor número de elementos en la muestra el error tiende a cero. Con esto hay que tener cuidado ya que está implícita la hipótesis de que la distribución de los estimadores es unimodal.

Otra forma de obtener los intervalos de confianza es realizando lo que se conoce como remuestreo (*bootstrapping*). Se considera a todo el conjunto de mediciones como la población completa y se toman muestras aleatorias de n elementos y se calcula el valor medio y la varianza de la muestra. El proceso se repite con reposición, es decir, nunca se retiran las mediciones del conjunto una vez que fueron elegidas de manera que pueden ser seleccionadas varias veces lo cual puede introducir un *bias* cuando el número de mediciones es pequeño (del orden de las decenas).

Lo siguiente a analizar es la distribución de longitudes de los pasos. Para eso se realiza un histograma de las mismas como se muestra en la figura 4.4, en este caso la presencia de máximos locales permiten identificar tamaños característicos en la longitud de los pasos. Por ejemplo, Coppin y colaboradores pudieron medir pasos nanométricos del motor molecular kinesina moviéndose sobre microtúbulos [Coppin et al., 1996]. Si la distribución de las componentes de los pasos en cada dirección es gaussiana entonces uno espera que la distribución de la longitud de los pasos estandarizados ($Z = (X - \mu) / \sigma$) siga la distribución de Rayleigh. Comparando con la función distribución obtenida estimando a partir de los datos el parámetro que la define, se puede evaluar si existe desviación de la misma.

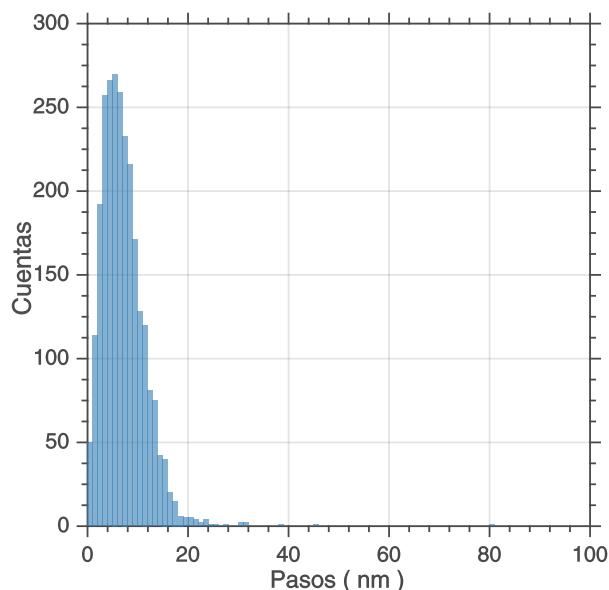


Figura 4.4: Histograma de las longitudes de los pasos presentados en la figura 4.3.

En el caso de que las distribuciones de las componentes de los pasos se asemejan a distribuciones gaussianas hay que evaluarlas para determinar si realmente pertenecen a esa familia de distribuciones. Esto aporta a la detección de más de una componente de movimiento en la trayectoria medida. Las distintas evaluaciones que se pueden realizar de forma gráfica son los q-q plots, normal plots, comparar el gráfico de la función distribución de probabilidad acumulativa empírica (ecdf) de la medición con la función distribución

acumulativa teórica utilizando como parámetros los valores estimados de las mediciones. Para dar un valor cuantitativo de si las mediciones perteneces a una distribución hay que realizar test estadísticos como lo son los test de Kolmogorov-Smirnov, Lilliefors y Anderson-Darling entre muchos otros, todos ya están implementados para los distintos programas como Matlab o Python. Previamente a utilizar los test es necesario revisar las hipótesis que tienen los mismos ya que es muy probable que se esté utilizando mal el test. Por ejemplo no se puede utilizar el test de Kolmogorov-Smirnov para comparar las mediciones contra la distribución con los parámetros estimados de los datos. Además hay que conocer la sensibilidad de los test a la presencia de *outliers* entre los datos. Por ejemplo en la medición que se muestra en la figura 4.3 si se evalúan los pasos descartando todos aquellos que tienen longitud mayor a tres desviaciones estándar σ el test de Lilliefors indica que se rechaza la hipótesis de que los pasos vienen de una distribución normal, pero si además se descartan los pocos pasos con longitud mayor a dos desviaciones estándar, el test da que no se rechaza la hipótesis de que los pasos provienen de una distribución gaussiana. Para más información sobre test de normalidad se puede consultar [Stephens, 1974]. Cuando la evaluación por una gaussiana da negativo se puede evaluar por movimiento con más de una componente.

Se puede suponer que las distintas componentes son gaussianas con distintos parámetros. Una complicación que se presenta en los sistemas con varias componentes es que la suma de dos distribuciones gaussianas $N(\mu_1, \sigma_1)$ y $N(\mu_2, \sigma_2)$, es unimodal si $|\mu_1 - \mu_2| \leq 2 \min(\sigma_1, \sigma_2)$ [Behboodian, 1970]. Lo cual hace que el problema de distinguir las distintas componentes sea complicado con herramientas usuales. Pero existen herramientas que permiten descomponer la distribución de las mediciones en suma de gaussianas como lo son los *Gaussian Mixture Models* (GMM). Para evaluar cuántas componentes tiene el movimiento se pueden usar tanto el AIC *Akaike Information Criterion* y el BIC *Bayesian Information Criterion* que penalizan la verosimilitud del modelo por el número de componentes usadas. Para más información sobre el uso de estos métodos se puede consultar [Hand et al., 1989].

El análisis de los pasos realizado hasta el momento se enfocó en el tamaño de los mismos, lo siguiente a analizar es la distribución angular de estos pasos, es decir, cuales son las direcciones (ángulo Θ respecto de un eje cartesiano) en las que apuntan los pasos y también cuales son los posibles cambios de dirección (ángulo ϕ entre dos pasos) que realiza la partícula (ver figura 4.5a).

Si el movimiento es isotrópico se espera que las direcciones de los pasos provengan de una distribución uniforme, es decir, no existe una dirección preferencial de movimiento. Y si los pasos son todos independientes entre ellos, la distribución de ángulos de giro también debe ser uniforme.

Para hacer un análisis estadístico de variables periódicas (ángulos) no se puede tomar el promedio como es usual. Por ejemplo, si se tienen dos mediciones de ángulos de 1° y 359° respecto del eje x , al promediarlos obtenemos que el ángulo medio es 180° , es decir, un vector apuntando en la dirección opuesta de los datos. Para solucionar esto, realizamos lo propuesto en [Scott, 2002, Berens, 2009, Zarr, 2010].

Primero se normaliza cada uno de los datos, de manera que todos los vectores pertenecen al círculo unitario. Suponiendo que se tienen N mediciones $\mathbf{X}_i = [x_i, y_i]$, se definen:

$$\begin{aligned} X &= \frac{1}{N} \sum_{i=1}^N x_i, & Y &= \frac{1}{N} \sum_{i=1}^N y_i, \\ R &= \sqrt{X^2 + Y^2}. \end{aligned} \tag{4.2}$$

A partir de los cuales se definen:

$$\cos \bar{\Theta} = \frac{X}{R}, \quad \sin \bar{\Theta} = \frac{Y}{R},$$

$$\theta_R = \arctan\left(\frac{\sin \bar{\Theta}}{\cos \bar{\Theta}}\right) = \arctan\left(\frac{Y}{X}\right). \quad (4.3)$$

Para analizar la distribución tanto de los ángulos direccionales como los ángulos de giro utilizar la función `plotAngle(walk)` que genera los gráficos de la derecha de la figura 4.5. En la figura se muestra los valores estadísticos junto con un histograma polar en azul, una muestra aleatoria de los pasos normalizados en verde y un vector en rojo que en dirección del ángulo promedio. La longitud del vector es el valor de R dado por la ecuación 4.2.

La distribución de los ángulos direccionales indica si hay correlación espacial y la distribución de ángulos de giro indica si existe correlación temporal entre los pasos. Lo que puede implicar que el movimiento sea dirigido. En el caso de que exista correlación ρ hay que corregir el coeficiente de difusión D por:

$$\tilde{D} = \frac{1+\rho}{1-\rho} D$$

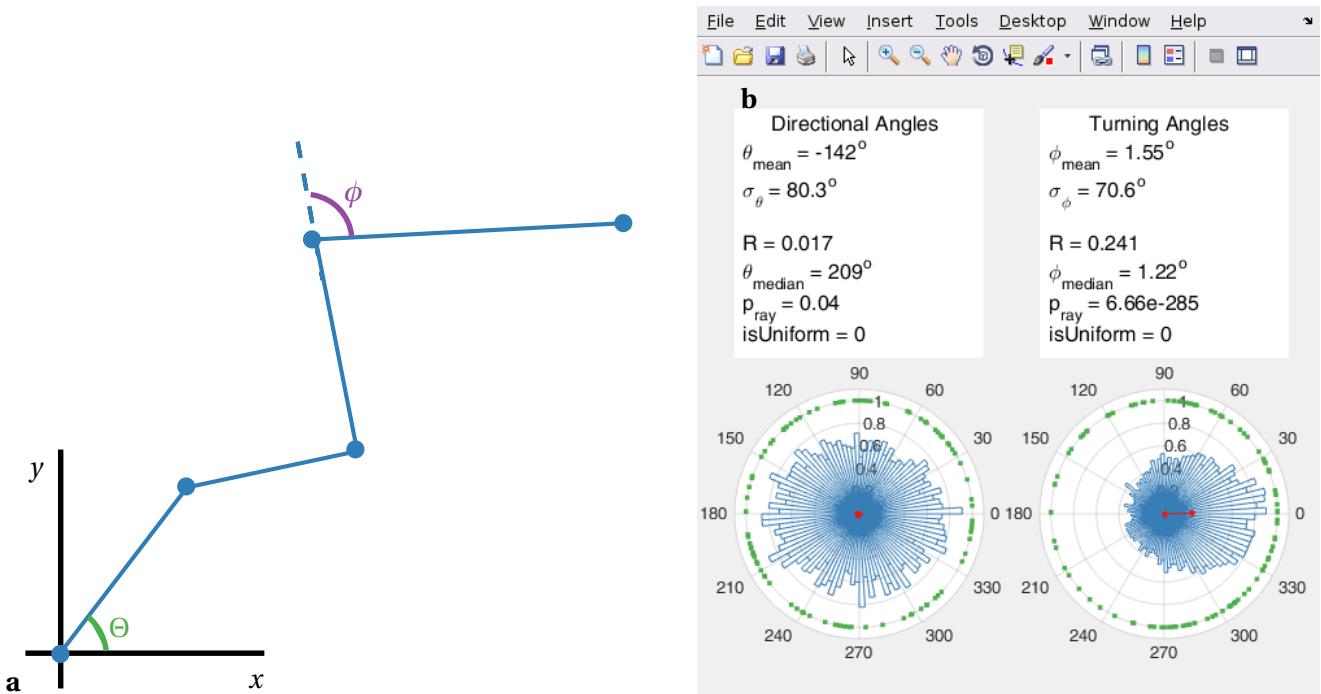


Figura 4.5: (a) Diagrama de los ángulos que forman los pasos de la caminata con los ejes cartesianos Φ (ángulos direccionales) y entre ellos ϕ (ángulos de giro). (b) Gráfico generado por la función `plotAngle()` donde se grafican histogramas polares de los ángulos direccionales y de giro. Los histogramas están escalados de manera que el radio del *bin* con más cuentas sea 1. En verde se grafican los pasos normalizados (en caso de que se tengan más de 100 pasos se grafica solo una muestra aleatoria de 100 elementos) y en rojo se grafica un vector en dirección del ángulo promedio y con longitud igual a R (4.2).

Para el análisis de las distribuciones angulares se puede usar el paquete de Circstat en el cual están implementados los test y adicionalmente la documentación del paquete da la descripción de los métodos [Berens

and Velasco, 2009]. Para más información de los distintos tests se puede consultar el *Handbook of Statistical Methods* de NIST [NIST, 2012, Spiegel et al., 2012, Hogg et al., 2014].¹

Para analizar la caminata se implementó en Matlab la clase `randomWalk()` con los métodos necesarios para graficar y calcular los distintos valores explicados hasta el momento. Ya analizados los pasos se pueden analizar las posiciones que toma la partícula. Para identificar regiones en las que la partícula pasa más tiempo graficar un histograma de las posiciones de la caminata obteniendo algo semejante a la figura 4.6. Los resultados que se obtienen de la segmentación pueden ser contrastados con lo obtenido en este histograma. Como complemento se puede ver la animación de la caminata (`walk`) utilizando la función `plotMovie(walk)`. Esto sirve para identificar puntos en los que se tiene que dividir la caminata.

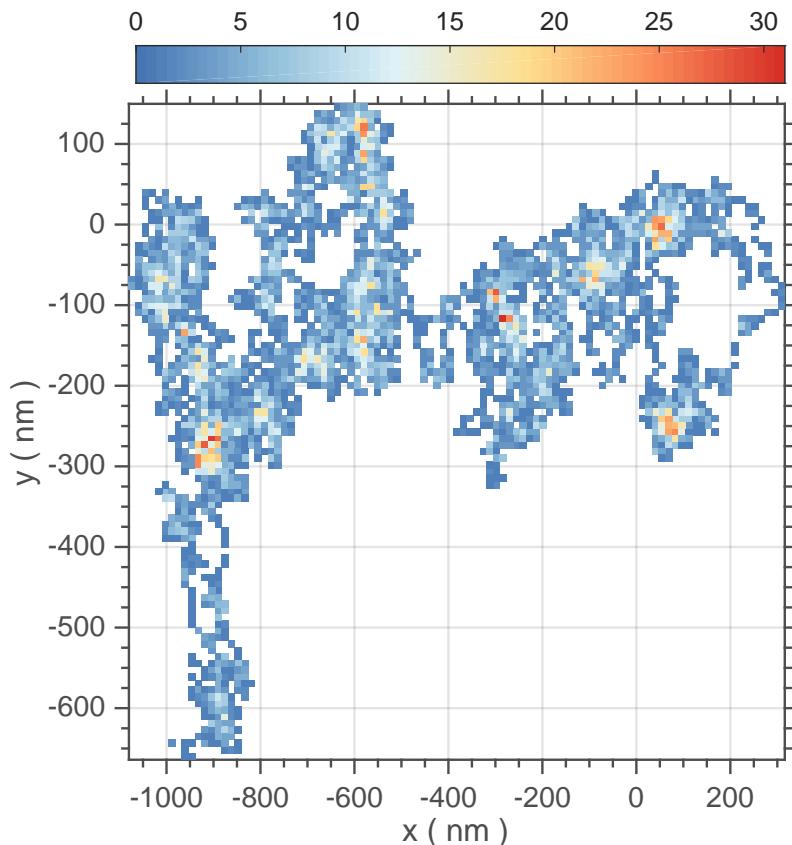


Figura 4.6: Histograma de una caminata experimental de 11000 pasos, los bins tienen un ancho de 13 nm.

Lo que resta para analizar es la presencia de distintos tipos de segmentos dentro de la trayectoria. Para lo cual se pueden utilizar las funciones `segmentThread(walk)` para calcular los filamentos. Para los cúmulos o regiones de confinamiento `segmentCluster(walk)` y para obtener los segmentos donde el movimiento tiene deriva `segmentByAnisotropy(walk)`. Adicionalmente, para observar los segmentos en la caminata utilizar `guiSegment(walk)` la cual es una interfaz gráfica en la que se puede seleccionar los distintos tipos de segmentos y da un reporte de la información referente a la caminata completa y al segmento seleccionado (fig. 4.7).

¹Para aquellos que no tienen conocimientos de probabilidad y estadística se recomienda el proyecto interactivo de la Universidad de Brown, Seeing Theory (enlace).

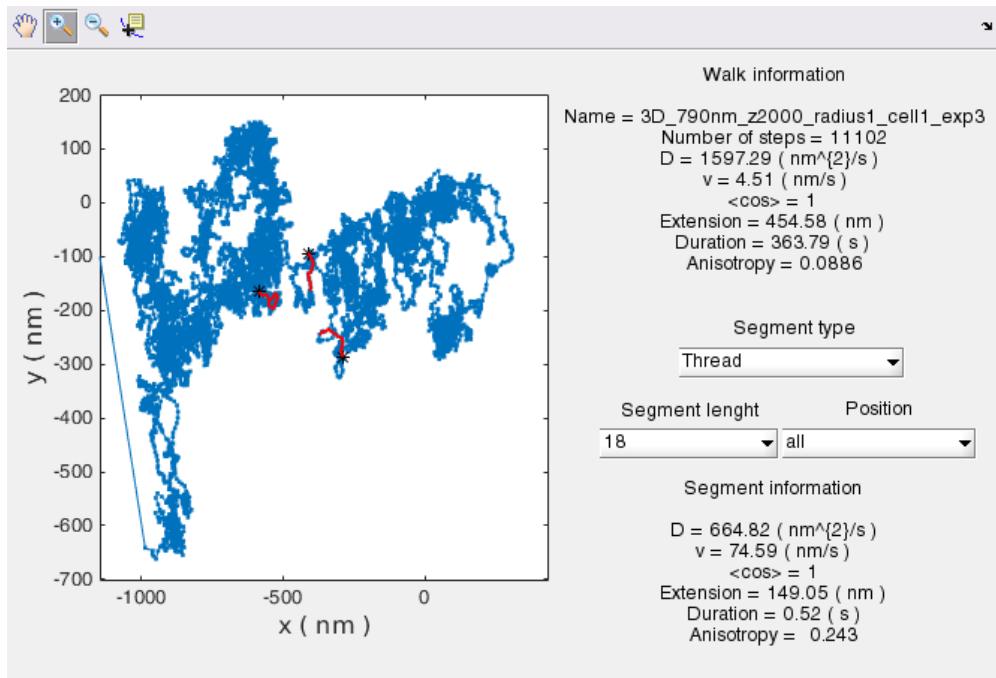


Figura 4.7: Captura de la ventana de la herramienta `guiSegment()` en la cual se muestra una caminata junto con tres filamentos (*threads*) de 18 pasos de duración (*Segment length*). Además se presenta la información general y valores de los distintos parámetros que describen a la caminata. Y en la parte inferior la información de los segmentos.

Incluimos una tabla de con los parámetros que se pueden evaluar en una caminata aleatoria.

Variable	Símbolo	Unidades
Valor medio de los pasos	μ	nm
Varianza de los pasos	σ^2	nm ²
Tiempo entre pasos	τ	s
Correlación	ρ	—
Ángulo direccional	Φ	rad/ ^o
Ángulo de giro	ϕ	rad/ ^o
Velocidad de deriva	ν	nm s ⁻¹
Coeficiente de difusión	D	nm ² s ⁻¹
coseno(ϕ) promedio	c	—
Anisotropía	a	—

Figura 4.8: Tabla de parámetros utilizados en el análisis de las trayectorias

PROTOCOLO

Distribución de los pasos •TIEMPO 5 min

Para determinar el coeficiente de difusión D_0 y la velocidad de deriva v utilizando la distribución de los pasos.

CUIDADO Revisar los datos en busca de valores singulares (*outliers*).

▲METODOS `plotScatter()`, `.v()`, `.D()`, `.meanStep()`, `.stepVariance()`

- 1| Realizar un gráfico polar de los pasos.
- 2| Calcular el valor medio de los pasos para cada dirección.
- 3| Evaluar si el valor medio es nulo utilizando un t-test.
- 4| Calcular la varianza de los pasos en cada dirección y la correlación.
- 5| Calcular el coeficiente de difusión y la velocidad de deriva.

Distribución angular de los pasos •TIEMPO 5 min

Para determinar la isotropía y homogeneidad del movimiento. Calcular el valor de la correlación espacial $\rho = c$ para corregir el coeficiente de difusión D

▲METODOS `plotAngle()`, `.meanCosine()`, `.stepAngle()`

- 6| Realizar histograma polar de los ángulos direccionales.
- 7| Calcular el ángulo medio Θ , su desviación estándar.
- 8| Calcular R y evaluar si los ángulos están distribuidos de forma uniforme.
- 9| Repetir los estos pasos para el ángulo de giro ϕ entre pasos.
- 10| Calcular el valor medio del coseno c y el seno s del ángulo de giro ϕ .
- 11| Corregir el coeficiente de difusión por $D = D_0(1 + c)/(1 - c)D_0$

PROBLEMA Verificar que el valor de s sea nulo, si no lo es la corrección del coeficiente de difusión se realiza por otra fórmula.

Segmentación de las caminatas •TIEMPO 10 min

▲METODOS `guiSegment()`, `segmentThread()`, `segmentCluster()`, `segmentByAnisotropy()`, `.XY()`

- 12| Realizar un histograma bivariado de las posiciones en las que estuvo la partícula.

PROBLEMA Se puede obtener un tamaño apropiado para los bins siguiendo la regla: tantos bins como la raíz cuadrada de la cantidad de pasos y un máximo de 30 bins.

- 13| Calcular los filamentos de la caminata.
- 14| Calcular los cúmulos de la caminata.
- 15| Calcular las regiones de movimiento dirigido.

TIEMPO El cálculo utilizando la anisotropía puede tomar un par de minutos en terminar. el tiempo crece de manera exponencial con la cantidad de pasos en la caminata.

Para información completa de los métodos y propiedades de `randomWalk()` ver el apéndice B

4. Análisis de trayectorias experimentales

5

Discusión y conclusiones

A woman asked,

Which way does the Wind Blow?

And he said:

The Wind is invisible but strong, like vodka...

The Wind is a friend.

The Wind is an enemy.

The Wind is neutral, like Switzerland...

Sometimes the Wind will blow down a tree.

Other times it won't.

There's just no telling with the Wind.

Which way will the Wind blow tomorrow, the woman persisted.

He answered:

The Wind is free and unpredictable, like a bird. It is impossible to guess which way it will blow.

She who would ask such a question would be a fool, and he who would attempt to answer it, a greater fool.

I understand, the woman said, but which way will it blow?

North, he replied.

The Profit

El trabajo se dividió en tres partes. Primero se desarrollaron algoritmos para simular caminatas aleatorias como procesos a tiempo discreto. Las caminatas se generaron bajo distintas condiciones: movimiento libre, en el cual el caminante tiene la misma probabilidad de moverse en cualquier dirección, caminatas confinadas en el espacio, caminatas con deriva generada por el movimiento del medio que transfiere una tendencia al caminante a moverse en una dirección preferencial y caminatas dirigidas generadas restringiendo el ángulo máximo entre 2 pasos consecutivos(caminata correlacionada en el tiempo). También con la condición de que la dirección de los pasos esté correlacionada en el tiempo, lo cual se genera restringiendo el ángulo máximo entre dos pasos consecutivos. Y como última condición, se generaron caminatas confinadas en el espacio.

Las simulaciones se utilizaron para estudiar las características del ensamble o conjunto de caminatas para los distintos tipos de movimiento. Y a partir de estos, se evaluaron los primeros dos momentos de la distribución de probabilidad de las posiciones que puede tomar el caminante en función del tiempo. Obteniendo que tienen una distinta forma funcional, dependiendo del tipo de movimiento.

Luego se analizaron las características del desplazamiento cuadrático medio (MSD) dado que es una herramienta utilizada frecuentemente en los experimentos de seguimiento de partícula única (SPT). El MSD es la suma de los dos primeros momentos de la distribución de probabilidad de las posiciones que puede tomar el caminante de manera que para calcularlo es necesario conocer la distribución de los mismos en función del tiempo. Si hay un único proceso generador de la caminata a lo largo de toda la trayectoria y cada paso es independiente del anterior (un sistema ergódico), es posible transformar el promedio sobre el ensamble a un promedio temporal. Esto se debe a que cada paso es independiente del tiempo, lo que permite que se

5. Discusión y conclusiones

pueda dividir una caminata en varias caminatas de menor duración y luego promediarlas entre sí. Cuando es posible suponer que la trayectoria experimental es ergódica el MSD se puede utilizar para distinguir el tipo de proceso que genera el movimiento, pero debe tenerse mucho cuidado si no es el caso.

En el análisis del MSD, se consideró al mismo como una función de la variable aleatoria posición del caminante, lo que implica que éste también es una variable aleatoria, es decir no es un valor fijo, sino que toma valores en un rango específico. Además este rango depende del tiempo en el cual se evalúa el MSD. De manera que los valores que toma el MSD calculado sobre una única caminata tienen un error intrínseco asociado a su origen estadístico. Este detalle es importante ya que es habitual que se clasifique una curva de MSD realizando un ajuste sobre la curva, y que el MSD tenga un error dependiente del tiempo obliga a que el ajuste se realice con valores pesados. En este trabajo analizamos la dependencia del error estadístico del MSD con el tiempo. Como complicación agregada, el MSD en el caso de caminatas libres sigue una distribución de chi-cuadrado, la cual es asimétrica, generando otro problema y otra suposición de los ajustes es que los valores que tome la variable estén distribuidos de manera simétrica alrededor del valor promedio. Entonces es más probable que una única curva de MSD esté por debajo del valor esperado.

Por lo expuesto, en esta tesis se analizó la validez de los resultados que se obtienen a partir de los ajustes del MSD. Se estudió si existe una dependencia del error de los parámetros estimados con el número de puntos de la curva del MSD que se utilizan para realizar el ajuste, obteniendo que la validez del resultado es independiente de la cantidad de puntos que se utiliza para ajustar el MSD si se utilizan ajustes pesados. Además la metodología permite que los resultados dependan del modelo que se utiliza para ajustar los datos.

Como indicamos, el análisis del MSD para una única caminata solo es válido si el sistema es ergódico. Esto no ocurre habitualmente en trayectorias experimentales, por ejemplo en el caso en el que una partícula (una vesícula o un motor molecular) se mueve en el interior celular provocando que una única caminata esté formada por porciones que obedecen a distintas dinámicas. Dado que los tiempos de interacción con los componentes de la célula son del orden del tiempo de medición. Para distinguir distintas porciones de la caminata tienen que ser tomadas en cuenta, se estudiaron criterios de selección binaria para clasificar entre caminatas libres y caminatas confinadas, con deriva o dirigidas. Luego se aplica este criterio a porciones de la caminata para segmentarla.

Está claro que como parte de una misma caminata libre es posible obtener porciones que correspondan a un movimiento dirigido. Para cuantificar este efecto se estudió la probabilidad de que en una caminata libre el ángulo entre dos pasos consecutivos se encuentre acotado, durante varios pasos consecutivos. Como resultado, se obtuvo que la probabilidad decrece exponencialmente con el número de pasos consecutivos de ocurrencia y con un exponente que depende del ángulo máximo que se considere. Este resultado implica que probabilidad de clasificar una caminata libre erróneamente como una dirigida decrece con la cantidad de pasos de la caminata.

Para el caso de confinamiento evaluamos la probabilidad de que una partícula se mantenga dentro de una región durante cierto tiempo y evaluamos esta probabilidad en función del tamaño de la región y el tiempo que permanece en la misma.

Para analizar el movimiento dirigido estudiamos la forma de la caminata evaluando la anisotropía de las posiciones que conforman la caminata. Se analizó la distribución de los valores que toma la anisotropía en función de la duración de la caminata y se obtuvo que ésta converge a una distribución a los pocos pasos en el caso de una caminata libre en la que los pasos están distribuidos de manera gaussiana. Cuando los pasos tienen otras características pero siguen teniendo valor medio nulo, la distribución tarda más en converger. Y en los casos de que el valor medio no sea nulo, las caminatas tienden a ser alargadas y la distribución de la anisotropía cambia en función del número de pasos de la caminata.

El criterio de selección de caminatas con deriva utiliza un valor umbral de anisotropía. Se estudió cuál es la calidad y poder del método en función del valor umbral de anisotropía. Para segmentar la caminata se realizó

un algoritmo en el cual se evalúa la anisotropía en todos los posibles fragmentos de la caminata, y para todos los fragmentos con distintos número de pasos. Como se evalúan segmentos con distinto número de pasos, hay segmentos que luego se seleccionan como parte de segmentos con más pasos. Con lo cual fue necesario generar un mapa en el que se describe la anisotropía en función del número de pasos del segmento y la posición del segmento para distinguir cuando ocurre esto. En el mapa se observan regiones con anisotropía similar las cuales se consideran como un único segmento, y para seleccionarlos se utiliza un algoritmo de segmentación de imágenes el cual separa las partes conexas, en este caso, regiones con anisotropía menor al valor umbral correspondiente con el 1-percentil de la distribución de anisotropía.

Se propuso un protocolo de análisis de caminatas aleatorias individuales basado en los contenidos del primer capítulo. Evaluando distintos parámetros de los pasos de la caminatas como lo son el valor medio y la varianza, además las distribuciones de los ángulos direccionales y los ángulos de giro.

Dentro del protocolo se presentan características que hay que evaluar y son de origen estadístico como lo son la independencia entre las direcciones de movimiento, y los intervalos de confianza para los parámetros estimados como el coeficiente de difusión y el valor medio de los pasos.

Para facilitar el uso de las herramientas aquí generadas se diseñó una aplicación en entorno Matlab que permite caracterizar cuantitativamente cualquier caminata independientemente de como haya sido generada. Esto no solo tiene aplicaciones en el estudio de la dinámica intracelular sino que también puede utilizarse para distinguir comportamientos singulares en sistemas no deterministas como el valor de una acción en el mercado financiero[Bachelier, 1900], el movimiento que realiza un animal en busca de alimento[Kareiva and Shigesada, 1983] o un comprador recorriendo un centro comercial. .

5. Discusión y conclusiones

A

Apéndices



A.1. Introducción de la Teoría de probabilidad

A.1.1. Procesos aleatorios

La aleatoriedad no es una propiedad de un solo número sino de una secuencia de números. Por ejemplo si hacemos un experimento en el que le preguntamos a alguien un número del uno al diez, sea cual sea la respuesta, no podemos saber si es una respuesta aleatoria porque no tenemos con qué compararla, puede ser que sea el número favorito de esa persona y siempre lo da como respuesta a esa pregunta, o que haya sido la primera vez que lo dijo. Por eso la aleatoriedad es la propiedad de una secuencia. Lo que hay que conocer, son los posibles resultados que se pueden obtener en un experimento, por ejemplo. Uno puede tener respuestas discretas como, si o no, o el resultado de tirar un dado, o incluso resultados que sean continuos, si es que se mide, longitud, masa Además, uno puede estar tentado a pensar que todas las respuestas que uno puede obtener salen con la misma frecuencia pero no tiene porqué ser así. Para que esto quede más claro. Si tiramos un dado, esperamos que cualquier número salga con la misma frecuencia (en el límite a muchas tiradas de dados). Pero si ahora tenemos dos dados. Hay una única forma de sacar 2 (sacando un 1 con cada dado), pero hay más formas de sacar un 7 ($1+6, 2+5, 3+4$). Conociendo esto, una secuencia aleatoria, es una secuencia en la cual conociendo los valores anteriores, no hay forma de predecir cuál va a ser el siguiente valor. Es decir, no hay forma de encontrar una estructura en la secuencia de datos. Veamos un par de ejemplos, si tomemos

varias secuencias de 10 números.

$$S_1 = 1111111111, \\ S_2 = 1212121212,$$

Ya conociendo estas secuencias, decimos que una secuencia es aleatoria si es que no se puede saber cuál va a ser el siguiente valor. Existen varias pruebas que se pueden realizar sobre una secuencia para determinar si la misma puede ser o no aleatoria. Otro aspecto importante en el desarrollo de esta tesis, es cómo se generan las secuencias aleatorias en una computadora. Estas secuencias no son exactamente aleatorias, sino, pseudo aleatorias porque uno puede generarla muchas veces.

Ya conociendo que significa que una secuencia sea aleatoria, pasemos a las definiciones y conceptos matemáticos utilizados en la tesis. La mayoría de las herramientas son conceptos básicos de probabilidad¹, pero la mayoría de los ejemplos se desarrollan en función de las caminatas aleatorias.

A.1.2. Introducción a la probabilidad

Uno introduce la noción de *evento*, los eventos genéricos los notamos por A, B, C, Pero un evento por sí solo no sirve de mucho, hay que contextualizar, cuando se define un modelo probabilístico, primero se define el *espacio muestral S*, que es el conjunto de todos los eventos posibles en el modelo, y el mismo también es considerado un evento en sí mismo. Entonces, la probabilidad es una regla que asigna un número real a cada evento, y usando la notación de $\Pr\{ \}$ para indicar la probabilidad del evento contenido entre las llaves, uno tiene que las reglas son:

1. $\Pr\{A\} \geq 0$ para todos los eventos A que pertenecen al espacio muestral S;
2. $\Pr\{S\} = 1$;
3. si A, B, C, \dots es una colección (posiblemente infinita) de eventos que son *mutuamente exclusivos* (es decir, no pueden ocurrir dos eventos en simultáneo),

$$\Pr\{A \text{ o } B \text{ o } C \text{ o } \dots\} = \Pr\{A\} + \Pr\{B\} + \Pr\{C\} \dots$$

Las primeras dos reglas marcan las convenciones de que las probabilidades son positivas y están normalizadas a uno, mientras que la tercera nos permite calcular la probabilidad de eventos compuestos en términos de la probabilidad de eventos más simples. Todas las demás reglas de la probabilidad se deducen de estas.

Por ejemplo, si notamos A' el evento *complementario* a A (es decir, el evento “A no ocurre”), entonces la regla (3) implica que

$$\Pr\{A\} + \Pr\{A'\} = 1. \quad (\text{A.1})$$

También se puede demostrar que si tenemos dos eventos que no son mutuamente exclusivos, entonces

$$\Pr\{A \text{ o } B\} = \Pr\{A\} + \Pr\{B\} - \Pr\{A \text{ y } B\}. \quad (\text{A.2})$$

¹Para visualizar todos los conceptos presentados, se puede utilizar la herramienta Seeing Theory (enlace) de la Universidad de Brown

La *probabilidad condicional* $\Pr\{A|B\}$ de un evento A , dado que el evento B haya ocurrido, se define por

$$\Pr\{A|B\} = \frac{\Pr\{A \text{ y } B\}}{\Pr\{B\}}, \text{ dado que } \Pr\{B\} \neq 0. \quad (\text{A.3})$$

Los eventos A y B se dicen *independientes* si y sólo si

$$\Pr\{A \text{ o } B\} = \Pr\{A\} \Pr\{B\}. \quad (\text{A.4})$$

A.1.3. Variable aleatoria

Una variable aleatoria X es una función la cual puede tomar distintos valores con sus respectivas frecuencias, cuando uno obtiene un valor de la función se denomina una realización de la misma y se nota como $X = x$. Lo que podemos saber de la variable aleatoria es su *mass density function* en el caso de que tome valores discretos o su *probability density function* en el caso de que tome valores continuos. Como son distribuciones de probabilidad la suma(integral) sobre todos los valores posibles tiene que dar 1.

Las variables aleatorias se definen como funciones que llevan eventos del espacio muestral hacia los números reales. Consideremos la variable aleatoria X definida sobre un espacio muestral discreto S , el cual tiene una secuencia numerable* de eventos elementales A_1, A_2, A_3, \dots en términos de las cuales todos los eventos en S pueden ser representados.

Con esto podemos pensar a X como una función en S , con la cual a cada evento elemental A_1, A_2, A_3, \dots le asociamos un número real $X(A_1), X(A_2), X(A_3), \dots$

Considerando todos los eventos A_i para los cuales $X(A_i) = x$, podemos definir de manera natural la probabilidad de que la variable aleatoria X tome el valor x :

$$\Pr\{X = x\} = \Pr\{A_1 \text{ o } A_2 \text{ o } A_3, \dots\}. \quad (\text{A.5})$$

Variables aleatorias en una dimensión

Una *variable aleatoria discreta* L tiene una *distribución de probabilidad* p_ℓ si

$$\Pr\{L = \ell\} = p_\ell \quad (\ell \text{ entero}), \quad (\text{A.6})$$

De la misma manera, una *variable aleatoria continua* X tiene una *densidad de probabilidad* $p(x)$ si

$$\Pr\{x < X < x + \delta x\} = \int_x^{x+\delta x} p(x') dx'. \quad (\text{A.7})$$

Y si x es un punto de continuidad de $p(x)$

$$\Pr\{x < X < x + \delta x\} \sim p(x)\delta x \quad \text{cuando } \delta x \rightarrow 0.$$

Las funciones p_ℓ y $p(x)$ tienen las siguientes restricciones

$$p_\ell \geq 0, \quad p(x) \geq 0,$$

y

$$\sum_{\ell=-\infty}^{\infty} p_{\ell} = 1, \quad \int_{-\infty}^{\infty} p(x) dx = 1.$$

Para unificar los casos de variables aleatorias discretas y variables aleatorias continuas es conveniente usar un formalismo más general que incluya ambos casos. El enfoque más simple utiliza la *delta de Dirac*, una función generalizada $\delta(x)$ con la propiedad

$$\int_{-\infty}^{\infty} \phi(x) \delta(x) dx = \phi(0), \quad (\text{A.8})$$

para todas las funciones suficientemente suaves $\phi(x)$.

La distribución de probabilidad p_l se corresponde con la densidad de probabilidad

$$p(x) = \sum_{\ell=-\infty}^{\infty} p_{\ell} \delta(x - \ell).$$

De manera vaga durante este trabajo va a ser suficiente con pensar a una variable aleatoria X simplemente como un *número indeterminado*, el cual puede asumir cualquiera de los valores de un conjunto. Con la elección de cada valor que pueda tomar X están asociadas probabilidades, como puede ser $\Pr\{X \geq 0\}$ o $\Pr\{-1 \leq X \leq 1\}$.

Variables aleatorias co-existentes

Podemos generalizar las ideas anteriores a variables aleatorias vectoriales que tienen densidades de probabilidad multidimensionales. De manera que la noción de *distribución conjunta* de varias variables aleatorias. Consideraremos el caso de dos variables aleatorias continuas X e Y . Si para alguna región Ω del plano tenemos que,

$$\Pr\{(X, Y) \in \Omega\} = \int \int_{\Omega} f(x, y) dx dy, \quad (\text{A.9})$$

entonces f es la densidad conjunta de X e Y . Las densidades de probabilidad

$$g(x) = \int_{-\infty}^{\infty} f(x, y) dy \quad y \quad h(y) = \int_{-\infty}^{\infty} f(x, y) dx \quad (\text{A.10})$$

propias de cada una de las variables aleatorias se llaman *densidades marginales* de X e Y respectivamente. Si dos variables aleatorias son consideradas QdeemedQ *independientes* si y sólo si su densidad de probabilidad conjunta se factoriza en el producto de las densidades marginales:

$$f(x, y) = g(x) h(y). \quad (\text{A.11})$$

Para dos variables aleatorias discretas, la definición de la distribución conjunta, de las distribuciones marginales y de independencia se obtienen reemplazando las integrales por sumas. Las definiciones para un

conjunto de n variables aleatorias pueden ser desarrolladas de la misma manera. Recalcamos que n variables aleatorias continuas (discretas) X_1, X_2, \dots, X_n son independientes si la densidad (distribución) conjunta toma la forma

$$f(x_1, x_2, \dots, x_n) = f_1(x_1)f_2(x_2)\cdots f_n(x_n), \quad (\text{A.12})$$

Con f_i la densidad (distribución) marginal de X_i .

Valores de expectación y momentos

El *valor medio, promedio o valor de expectación* de cualquier función $f(X)$ de una variable aleatoria X (con densidad $p(x)$) está definido por

$$\langle f(X) \rangle = \int_{-\infty}^{\infty} f(x)p(x)dx \quad (\text{A.13})$$

Hay funciones que los valores de expectación de funciones simples tienen mayor significado físico o son más fáciles relacionarlos con que la distribución o densidad en su totalidad. Son de particular importancia los *momentos* de orden μ :

$$\langle L^\mu \rangle = \sum_{\ell=-\infty}^{\infty} \ell^\mu p_\ell \quad \text{y} \quad \langle X^\mu \rangle = \int_{-\infty}^{\infty} x^\mu p(x)dx. \quad (\text{A.14})$$

Los momentos de mayor importancia son, el momento de orden 1, el *promedio usual*, y el momento de orden 2 que denominaremos *promedio cuadrático*. También se utilizan los momentos centrados definidos por $\langle [X - \langle X \rangle]^\mu \rangle$. El momento central de orden 2, denominado *varianza* está definido por

$$var(X) = \langle [X - \langle X \rangle]^2 \rangle = \langle X^2 \rangle - \langle X \rangle^2. \quad (\text{A.15})$$

Como $var(X) \geq 0$, se tiene que $\langle X^2 \rangle \geq \langle X \rangle^2$, y la igualdad vale en el caso de que la variable X tome un 'único' valor.

Resultados importantes

Presentamos un par de resultados importantes que nos van a ser de utilidad.

DESIGUALDAD DE CHEBYSHEV. Para una variable aleatoria X con *valor medio finito* μ y *varianza finita* σ^2

$$\Pr\{|X - \mu| \geq k\sigma\} \leq \frac{1}{k^2}. \quad (\text{A.16})$$

DESIGUALDAD DE JENSEN. Tanto para variables aleatorias continuas como discretas, siempre que los valores que puede tomar la variable aleatoria X estén en un intervalo en el que la función f es convexa,

$$\langle f(X) \rangle \geq f(\langle X \rangle). \quad (\text{A.17})$$

DESIGUALDAD DE CAUCHY-BUNYAKOVSKY-SCHWARZ.

$$\langle XY \rangle^2 \leq \langle X^2 \rangle \langle Y^2 \rangle. \quad (\text{A.18})$$

TEOREMA CENTRAL DEL LÍMITE La suma de un número grande de variables aleatorias independientes e idénticamente distribuidas con varianza finita, puede aproximarse por una variable aleatoria con distribución gaussiana o normal. Para ver todos los detalles sobre el teorema se puede consultar [Gnedenko and Kolmogorov, 1954, Feller, 1968]

A.1.4. Operaciones con variables aleatorias

El $\langle X \rangle$ es el valor medio de la función y es un promedio de todas las posibles realizaciones de la variable aleatoria, en matemática se denomina el valor de expectación de la variable. El $var(X)$ determina la distancia promedio de los valores que puede tomar la distribución alrededor del valor medio y se denomina varianza.

Vamos a ver que con estas definiciones nos basta para estudiar el comportamiento de las caminatas aleatorias.

Por ahora solo hablamos de tener una única variable aleatoria, pero se puede analizar el caso en el que se tenga más de una. Regresando al caso de los dos dados. Conocemos la distribución de probabilidad de cada uno, ahora supongamos que nuestro experimento consiste en tirar ambos dados y considerar la suma de los valores que salen. Es decir, decimos que X e Y son las variables aleatorias que representan a los dados, cada una con su distribución de probabilidad $p_X(x)$ y $p_Y(y)$ ahora definimos una nueva variable aleatoria que sea la suma de ambas $Z = X + Y$ nos interesa saber cual es la distribución de probabilidad de Z

$$p_Z(z) = \int_{-\infty}^{\infty} p_{XY}(z-y, y) dy = \int_{-\infty}^{\infty} p_{XY}(x, z-x) dx \quad (\text{A.19})$$

En el caso de que las variables sean independientes,

$$p_Z(z) = \int_{-\infty}^{\infty} p_X(z-y) p_Y(y) dy = \int_{-\infty}^{\infty} p_X(x) p_Y(z-x) dx \quad (\text{A.20})$$

Lo que estamos haciendo es considerar la probabilidad de tener una variable Z con el valor z ver cuales son las combinaciones de x e y que permiten este resultado.

Para calcular los valores medios tenemos que

$$\langle X + Y \rangle = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x + y) p_{XY}(x, y) dx dy = \langle X \rangle + \langle Y \rangle, \quad (\text{A.21})$$

Algo que tenemos que definir es la probabilidad conjunta de X e Y , $p_{XY}(x, y)$, que tiene las siguientes propiedades. int sobre X, Y da las marginales de cada una.

Hay que definir la independencia entre dos variables, y eso se da si la En el caso de que X e Y sean independientes, tenemos que la probabilidad conjunta se factoriza. Y la varianza en el caso más general es

Para el caso de la varianza tenemos que

$$var(X + Y) = var(X) + var(Y) + 2 cov(X, Y), \quad (\text{A.22})$$

Podemos generalizar estos resultados para la suma de m'as variables aleatorias.

$$\left\langle \sum_{i=1}^N X_i \right\rangle = \sum_{i=1}^N \langle X_i \rangle \quad (\text{A.23})$$

$$\text{var}\left(\sum_{i=1}^N X_i\right) = \sum_{i=1}^N \text{var}(X_i) + 2 \sum_{i=1}^N \sum_{j \neq i}^N \text{cov}(X_i, X_j), \quad (\text{A.24})$$

La covarianza tiene las siguientes propiedades, si X e Y son independientes, su valor es nulo. En este caso la varianza de la suma se reduce a la suma de las varianzas $\text{cov}(X, Y) \leq \text{var}(X) \text{var}(Y)$

A.2. Caminatas Aleatorias

Una caminata aleatoria se define como la trayectoria en la que cada paso se realiza de manera aleatoria, y con todos los pasos idénticamente distribuidos, con esto nos referimos a que todos los pasos tienen la misma distribución de probabilidad.

Sean X_k un conjunto de variables aleatorias idénticamente distribuidas, se define la caminata aleatoria S_n como:

$$S_n = X_1 + X_2 + \cdots + X_n = \sum_{i=1}^n X_i, \quad (\text{A.25})$$

Podemos generalizar los resultados A.21 y A.22 a la suma de n variables, obteniendo que el valor medio y la varianza de una caminata aleatoria está dada por

$$\langle S_n \rangle = \sum_{i=1}^n \langle X_i \rangle, \quad (\text{A.26})$$

$$\text{var}(S_n) = \sum_{i=1}^n \text{var}(X_i) + 2 \sum_{i=1}^n \sum_{j \neq i}^n \text{cov}(X_i, X_j) \quad (\text{A.27})$$

Que en el caso de una caminata aleatoria donde las variables idénticamente distribuidas, entonces tienen el mismo valor medio y la misma varianza, y son independientes entonces la covarianza es nula para todo $i \neq j$

$$\langle S_n \rangle = n \langle X \rangle, \quad (\text{A.28})$$

$$\text{var}(S_n) = n \text{var}(X) \quad (\text{A.29})$$

Es importante notar que en todo este desarrollo no hicimos ninguna suposición sobre el espacio en el que se realizan las caminatas ni si la distribución de pasos es discreta o continua. A continuación vamos a estudiar varios casos en una dimensión, más adelante veremos cómo extendemos estos resultados a más dimensiones

A.2.1. Caminatas libres

El caso más sencillo de una caminata aleatoria es ocurre cuando los pasos tienen valor medio nulo, de forma que la caminata también tiene valor medio nulo.

Supongamos que tenemos una caminata en 1 dimensión donde la longitud de los pasos es constante (ℓ) y hay la misma probabilidad de avanzar que de retroceder entonces la distribución de probabilidad de los pasos es

$$p_j = \frac{1}{2} \quad \text{si} \quad |X_j| = \ell, \quad (\text{A.30})$$

ahora podemos estudiar la probabilidad de la caminata a medida que da más pasos, podemos utilizar la fórmula A.20 , obteniendo

$$p(m, n) = \left(\frac{1}{2}\right)^n \binom{n}{\frac{n-m}{2}}. \quad (\text{A.31})$$

que es la distribución binomial que nos dice cual es la probabilidad de encontrar a la partícula en la posición k si la caminata tiene n pasos

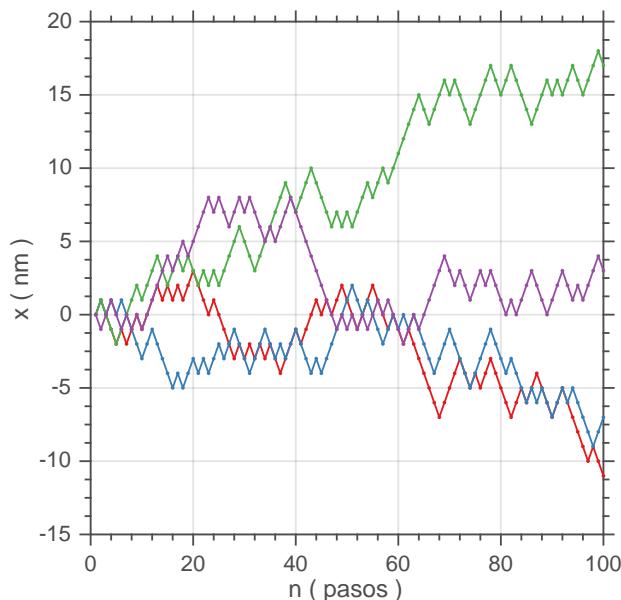


Figura A.1: Caminata unidimensional sobre un *lattice*, los pasos tienen longitud constante $\ell = 1 nm$. Todas las simulaciones tienen una duración de 100 pasos.

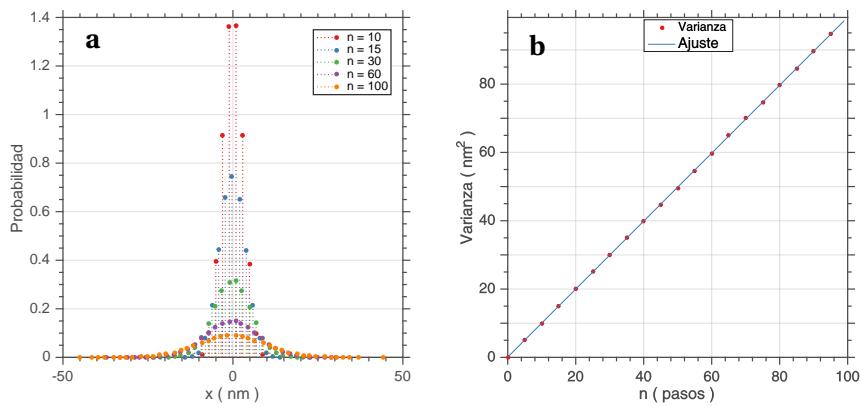


Figura A.2: (a) Histograma de la probabilidad de encontrar un caminante a una distancia x del origen. (b) Varianza de las posiciones finales después de n pasos. En azul se muestra el ajuste por una recta con $m = 0,990 \pm 0,002$ y $b = 0,09 \pm 0,10$, $R^2 = 1,0$

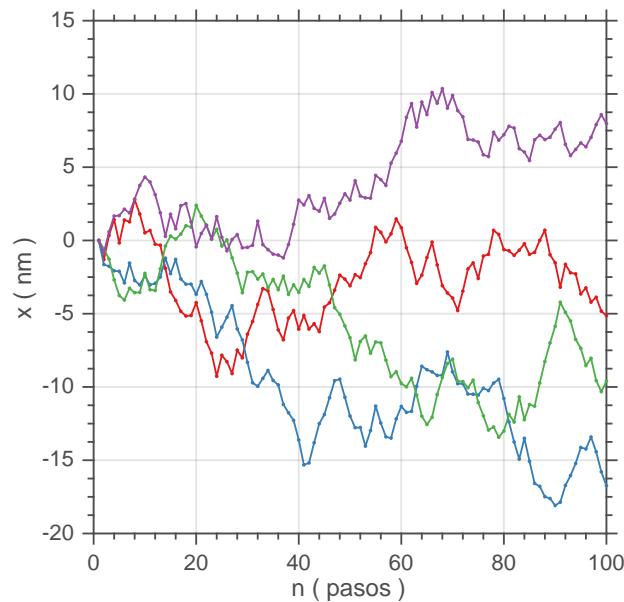


Figura A.3: Caminatas aleatorias con distribución de pasos uniforme con varianza $\sigma^2 = 1 \text{ nm}^2$.

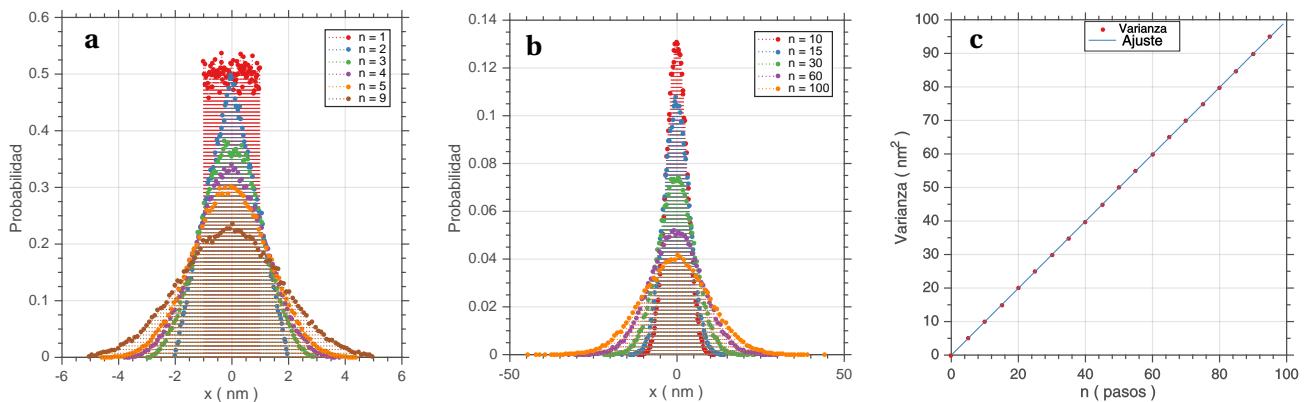


Figura A.4: (a) y (b) Histograma de la probabilidad de encontrar un caminante a una distancia x del origen. (c) Varianza de las posiciones finales después de n pasos. En azul se muestra el ajuste por una recta con $m = 1 \pm 1,3 \times 10^{-3}$ y $b = -0,03 \pm 0,07$, $R^2 = 1,0$

Ahora consideremos lo siguiente, supongamos que tenemos un experimento en el que cada paso se realiza cada τ segundos de manera que podemos definir el tiempo de la caminata como $t = n \tau$ y definimos una distancia característica a partir de la varianza de los pasos $\text{var}(X) = \ell^2$ entonces definimos el coeficiente de difusión como²

$$D = \frac{\text{var}(X)}{2 \tau} = \frac{k_b T}{6\pi\mu R_p}, \quad (\text{A.32})$$

La última igualdad es el resultado de Einstein[Einstein, 1905] para una partícula en un medio *isotrópico*. Las constantes son: k_b es la constante de Boltzman, T es la temperatura absoluta, μ es el coeficiente de viscosidad y R_p es el radio de la partícula. Recalcamos que este resultado es para medios *isotrópicos*. Para el caso de una partícula moviéndose en una membrana ver el trabajo de [Saffman and Delbrück, 1975].

El dos en el denominador es una convención para simplificar las cuentas cuando se estudia la ecuación de difusión.

Para generar la caminata en la computadora realizamos algo como esto, aprovechando que *Matlab* es bueno para vectorizar operaciones³.

Con esto ya podemos generar las caminatas que sean necesarias

Veamos el cálculo explícito del valor medio y la varianza de los pasos

$$\begin{aligned} \langle X \rangle &= \sum x p(x) = \frac{1}{2}(-\ell) + \frac{1}{2}\ell = 0, \\ \text{var}(X) &= \sum (x - \langle X \rangle)^2 p(x) = \frac{1}{2}(-\ell)^2 + \frac{1}{2}(\ell)^2 = \ell^2 \end{aligned}$$

De manera que para la caminata tenemos que

²Esto viene motivado por el caso en el que los pasos tengan un valor constante ℓ y la caminata sea simétrica, es decir hay la misma probabilidad de avanzar que retroceder, con lo que la varianza es igual a la longitud del paso al cuadrado.

³Vectorizar una operación quiere decir que se puede realizar una misma operación sobre toda un vector sin necesidad de recurrir a una iteración

$$\langle S_n \rangle = 0, \quad var(S_n) = n\ell^2.$$

Podemos observar en los gráficos como es el comportamiento de la distribución de probabilidad de las caminatas, para hacerlo un poco más tangible, pensemos que a tiempo $n = 0$ se tiene un gran número de partículas que no interactúan entre ellas y cada una realiza una caminata aleatoria con paso fijo ℓ , se liberan todas al mismo tiempo y se ve un tiempo posterior la cantidad de partículas que hay a cada distancia, a medida que pasa el tiempo, las partículas se van alejando de la posición de origen dispersando cada vez más. Por lo que se observa que la distribución de probabilidad se achata cada vez más y al mismo tiempo se ensancha indicando que las partículas recorrieron más distancia.

Esto puede parecer curioso de que solo importe la varianza de la distribución para conocer el comportamiento, pero hay que tener en cuenta que uno está analizando valores medios. Algo más extraño es que la distribución de la suma tiende a una normal, esto es lo que se conoce como el Teorema Central del Límite [Gnedenko and Kolmogorov, 1954].

Ahora veamos el caso en el que la distribución de probabilidad de los pasos tiene una probabilidad $p/2$ de moverse (avanzar o retroceder) $p \leq 1$ y una probabilidad $(1-p)$ de quedarse quieta.

El valor medio de los pasos sigue siendo nulo,

$$\langle X \rangle = 0, \tag{A.33}$$

Y la varianza es

$$var(X) = \frac{p}{2}(-\ell)^2 + (1-p)0 + \frac{p}{2}(\ell)^2, = p\ell^2,$$

de manera que ahora la varianza de la caminata es m

$$var(S_n) = n p \ell^2 \tag{A.34}$$

la cual es menor a la del caso anterior.

Esto es, si una partícula tiene la posibilidad de quedarse quieta, en el transcurso del tiempo va a recorrer menos distancia, de manera que la varianza de los pasos nos da una idea de la distancia que puede recorrer la partícula en cierto tiempo.

Si ahora vemos el caso el que la partícula puede dar pasos de distintas longitudes y además quedarse quieta, supongamos que tiene la misma probabilidad de ocurrir. Supongamos que puede dar pasos de longitud $0, \ell, 2\ell, 3\ell, \dots, r\ell$ de manera que hay la probabilidad de cada paso es $1/(2r+1)$.

El valor medio de los pasos es nulo y la varianza es:

$$\begin{aligned}
var(X) &= \sum_{i=1}^{2r+1} \frac{1}{2r+1} x_i^2, \\
&= \sum_{i=-r}^r \frac{1}{2r+1} (il)^2, \\
&= \frac{\ell^2}{2r+1} \left[2 * \frac{1}{6} r(r+1)(2r+1) \right] \\
&= \frac{1}{3} \ell^2 r(r+1) \\
var(S_n) &= \frac{r(r+1)}{3} n \ell^2
\end{aligned}
\tag{A.35}$$

$$var(S_n) = \frac{r(r+1)}{3} n \ell^2 \tag{A.36}$$

Con lo que tenemos un proceso en el que la distancia media recorrida es menor que si tenemos una caminata de longitud de paso único.

Con la misma idea podemos tener una caminata en la que los pasos tienen una distribución continua y los resultados son exactamente los mismos, solo hay que cambiar las sumatorias por integrales.

Poner gráficos de las distribuciones de probabilidad para distintos pasos

Hay algo que notar. En los ejemplos anteriores, tomamos las distribuciones de pasos uniformes, pero eso no tiene porqué ser así. Y hay que observar algo. Siempre que tenemos una caminata con pasos de varianza finita y valor medio nulo, tenemos que a medida que los pasos aumentan, las distribuciones de pasos siempre tienden a ser más acampanadas. Esto es un resultado general que está relacionado al teorema central del límite.

Supongamos que tenemos una caminata de N pasos lo suficientemente larga para poder extraer una subcaminata $\hat{S}_k = S_{r*k}$, es decir, se tiene menor resolución temporal sobre la caminata. Si el la nueva caminata observamos la distribución de pasos, lo que tenemos es que los pasos tienen la distribución de una caminata de r pasos. De manera que si no se tiene la suficiente resolución uno va a observar una distribución acampanada.

Con lo que obtenemos que a partir de cierto número de pasos todas las caminatas aleatorias se comportan de la misma manera, y lo único que interesa es la varianza de los pasos.

En el caso de tener una distribución de pasos discreta, la aproximación es válida cuando la varianza es mucho menor a la escala que se está considerando en el problema de manera que el espaciado de los pasos es infinitesimal en comparación de las distancias recorridas.

Límite de caminata 1d

Como el problema es discreto, en vez de considerar la probabilidad $P(x_0|x; t)$ utilizamos $P(n\Delta|m\Delta; s\tau) = P(n|m; s)$ que es la probabilidad de que la partícula esté en $m\Delta$ en el tiempo $s\tau$ si al principio se encontraba en la posición $n\Delta$.

La probabilidad de encontrar a la partícula en la posición $m\Delta$ en el paso $s+1$ depende solamente de las probabilidades de encontrarla en las posiciones vecinas en el paso anterior s , es decir, el proceso es Markoviano y $P(n|m, s)$ cumple:

$$P(n|m; s+1) = \frac{1}{2} P(n|m-1; s) + \frac{1}{2} P(n|m+1; s) \tag{A.37}$$

uno puede reescribir de manera equivalente como:

$$\frac{P(n\Delta|m\Delta;(s+1)\tau) - P(n\Delta|m\Delta;s\tau)}{\tau} = \frac{\Delta^2}{2\tau} \left\{ \frac{P(n\Delta|(m+1)\Delta;s\tau) - 2P(n\Delta|m\Delta;s\tau) + P(n\Delta|(m-1)\Delta;s\tau)}{\Delta^2} \right\} \quad (\text{A.38})$$

Considerando el límite donde Δ y τ se aproximan a 0 de forma tal que:

$$\frac{\Delta^2}{2\tau} = D, \quad n\Delta \rightarrow x_0, \quad s\tau \rightarrow t; \quad (\text{A.39})$$

Esta es una forma poco rigurosa de llegar a la ecuación de difusión para la probabilidad, pero da una idea clara de cómo se relacionan ambas descripciones.

Una manera de estudiar las caminatas aleatorias es mediante la probabilidad

$$\int_{x_1}^{x_2} P(x_0|x;t) dx, \quad (\text{A.40})$$

de que a un tiempo t una partícula se encuentre entre las posiciones x_1 y x_2 si estuvo en la posición x_0 al tiempo $t=0$.

Einstein[Einstein, 1905] consideró el caso de una partícula que se mueve libremente en un medio homogéneo, lo que se conoce como Movimiento Browniano, y probó que la probabilidad para este caso tiene que cumplir con la ecuación de difusión

$$\frac{\partial P}{\partial t} = D \frac{\partial^2 P}{\partial x^2}, \quad (\text{A.41})$$

la cual tiene como solución:

$$P(x_0|x;t) = \frac{1}{\sqrt{4\pi D t}} \exp\left[-\frac{(x-x_0)^2}{4Dt}\right], \quad (\text{A.42})$$

Esta solución solo se puede considerar para tiempos largos, es decir, para procesos en los que hay un gran número de pasos.

El movimiento Browniano es el caso más sencillo de una caminata aleatoria, en esta tesis estudiamos una variedad de casos como caminatas en las que existe una dirección en la que la partícula tiene una preferencia para moverse, caminatas sobre un flujo, es decir existe una velocidad adicional constante. Caminatas que están confinadas y que tienen cierta probabilidad de escapar del confinamiento entre otros casos. A continuación presentamos los distintos casos realizados mediante simulaciones numéricas junto con el cálculo de varios parámetros que nos permiten caracterizar cada tipo de caminata lo cual utilizaremos luego para analizar las trayectorias experimentales.

A.2.2. Caminatas con deriva

Ahora podemos considerar el caso en el que el valor medio de los pasos no es nulo. Lo que esto implica es que el valor medio de la caminata evoluciona con el tiempo.

Esto puede ocurrir ya sea porque existe alguna preferencia para el movimiento, no es un medio isótropo (existe un gradiente de concentración) o existe un flujo constante de manera que ahora la variable aleatoria se puede pensar como

$$\langle X \rangle = \ell_0 \neq 0.$$

Con esto podemos considerar que tenemos cubierta lo más importante de las caminatas aleatorias, en una sola dimensión y podemos pasar a más dimensiones.

Vamos a ver el comportamiento de las caminatas aleatorias cuando el valor medio de los pasos no es nulo, y ver que es equivalente a tener un sistema en el cual la partícula se encuentra en un fluido con velocidad constante

Teníamos que el valor medio de la caminata A.27 era proporcional al valor medio de los pasos. Podemos reescribirla definiendo

$$\langle S_n \rangle = n \ell_0, \quad (\text{A.43})$$

$$\text{var}(S_n) = n \text{var}(X) \quad (\text{A.44})$$

Al igual que se definió el coeficiente de difusión en función del tiempo característico τ , definimos la velocidad de deriva como

$$v = \frac{\ell_0}{\tau}. \quad (\text{A.45})$$

Esto lo podemos tomar como el equivalente a tener un término constante en la ecuación que define la caminata aleatoria

$$S_k = S_{k-1} + x_k + v, \quad (\text{A.46})$$

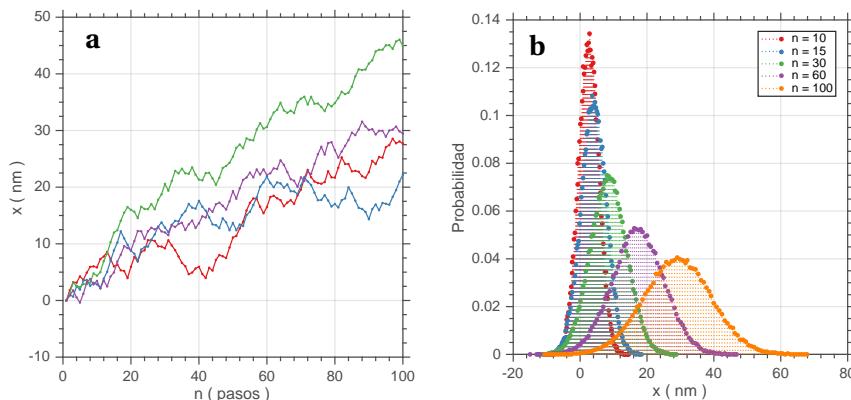


Figura A.5: Caminatas unidimensionales con distribución de pasos uniformes de varianza $\sigma^2 = 1 \text{ nm}^2$ y valor medio $\ell_0 = 0,3 \text{ nm}$

En el caso de más dimensiones se puede tratar similar de varias formas, ya sea se puede colocar un flujo que

hace que la partícula convecte o utilizar distribuciones ya sea de longitud de pasos o angular que tengan con valor medio distinto de cero.

Cuando en el sistema existe una fuerza, o una dirección preferencial de movimiento es necesario modificar la ecuación A.41. Smoluchowski presentó la generalización en la que aparece aparece un término de flujo.

$$\frac{\partial P}{\partial t} = D \frac{\partial^2 P}{\partial x^2} - \frac{1}{f} \frac{\partial}{\partial x} (P F), \quad (\text{A.47})$$

Esta ecuación tiene como solución más sencilla,

$$P(x_0|x; t) = \frac{1}{\sqrt{4\pi D t}} \exp\left[-\frac{(x - vt)^2}{4Dt}\right], \quad (\text{A.48})$$

Veamos el caso unidimensional discreto. Se puede pensar que la probabilidad de moverse en una dirección o en la otra ya no es simétrica. Suponiendo que la probabilidad de avanzar es $p = 1/2 - \beta\Delta$, por lo tanto la probabilidad de retroceder es $q = 1/2 + \beta\Delta$ donde β es una constante física a determinar en los experimentos, Δ debe ser un valor pequeño tal que $p > 0$.

$$P(n|m; s+1) = p P(n|m-1; s) + q P(n|m+1; s), \quad (\text{A.49})$$

De manera similar a lo realizado para la probabilidad de la partícula libre, se puede reescribir esta última ecuación y tomando los límites A.39 se llega a la forma:

$$\frac{\partial P}{\partial t} = D \frac{\partial^2 P}{\partial x^2} + 4\beta D \frac{\partial P}{\partial x}, \quad (\text{A.50})$$

que es de la forma de A.47 con $F(x) = -a = -4\beta D f$, una partícula con un deriva constante. Y la generalización a más dimensiones es [Smoluchowski, 1976]

$$\frac{\partial P}{\partial t} = \nabla \cdot (\mathbf{D} \nabla P) - \nabla(\vec{u} P), \quad (\text{A.51})$$

En esta ecuación \mathbf{D} es un tensor de manera que la partícula puede difundir con distintas velocidades según la dirección de movimiento.

[Rycroft and Bazant, 2005]

A.2.3. Caminatas en d dimensiones

Para analizar caminatas aleatorias en más dimensiones, supongamos que tenemos una variable aleatoria $\mathbf{X} = [X_1, X_2, \dots, X_d]$ que ahora pertenece a un espacio vectorial de d dimensiones.

Ahora la distribución de probabilidad de $p(\mathbf{X})$ es la distribución conjunta del conjunto de d variables aleatorias $p(X_1 = x_1, \dots, X_d = x_d)$.

Con esto podemos calcular el valor medio de la caminata, con lo que podemos observar que el valor medio de

$$\langle \mathbf{X} \rangle = [\langle X_1 \rangle, \langle X_2 \rangle, \dots, \langle X_d \rangle] \quad (\text{A.52})$$

es el vector en el cual se tiene el valor medio de cada una de las componentes.

Ahora en el caso de la varianza, tenemos que recordar que $\mathbf{X}^2 = X_1^2 + X_2^2 + \dots + X_d^2 = R^2$, lo que nos lleva a

$$\text{var}(\mathbf{X}) = \sum_{i=1}^d \text{var}(X_i) + 2 \sum_{i=1}^d \sum_{j \neq i} \text{cov}(X_i, X_j) \quad (\text{A.53})$$

que en el caso de que todas las componentes de movimiento sean independientes, de forma que la covarianza es nula. Y si además los pasos en cada dirección están idénticamente distribuidos, tenemos que la varianza.

Si cada una de las componentes es independiente entonces,

$$p(X_1 = x_1, X_2 = x_2, \dots, X_d = x_d) = \prod_{i=1}^d p(X_i = x_i) \quad (\text{A.54})$$

En el caso de movimiento independiente en cada dirección y idénticamente distribuido (movimiento isotrópico), la ecuación A.53 se reduce a:

$$\begin{aligned} \langle R^2 \rangle &= \text{var}(\mathbf{X}) = d \text{var}(X) \\ &= 2dD t, \end{aligned} \quad (\text{A.55})$$

Un caso interesante es el de tener movimiento en varias dimensiones, pero permitir que cada paso sea dado en una única dimensión.

Entonces tenemos que

$$p(X_i | X_j) = \begin{cases} 1 & \text{si } X_j = 0 \quad \forall j \neq i, \\ 0 & \text{en caso contrario} \end{cases} \quad (\text{A.56})$$

Hacer esta distinción es importante en el momento de realizar las simulaciones de las caminatas aleatorias, uno podría estar interesado en el caso de caminatas en las que la longitud del paso es constante (que por facilidad conviene fijar en 1) y una forma de realizarlo es pensar el movimiento sobre \mathbb{Z}^d o en \mathbb{R}^d normalizando cada paso lo que hace que las coordenadas dejen de ser independientes.

Veamos como sería el código para una caminata en varias dimensiones

En los casos multidimensionales lo que es de interés es obtener la distribución de probabilidad de la longitud de los pasos en función de las distribuciones en cada una de las coordenadas. Algo que hay que notar que esto es necesario ya que en un experimento en el que la partícula se mueve en varias dimensiones se pierde la noción de si la partícula avanza o retrocede en cada paso.

Para calcular esto hay que tener en cuenta esta igualdad

$$p(X_1 = x_1, \dots, X_d = x_d) dx_1 dx_2 = p(R, \phi_1, \dots, \phi_{d-1}) dR d\phi_1 \cdots d\phi_{d-1}, \quad (\text{A.57})$$

Completar con los resultados para las distribuciones de probabilidad y dar un apéndice de las relaciones entre las distintas.

En el caso 2d hacer un par de caminatas y mostrar cuales son los resultados. Y también hacer un par de caminatas en 3d.

A.2.4. Caminatas confinadas

Hasta ahora consideramos caminatas que pueden recorrer todo el espacio, pero es posible restringir esto a regiones acotadas S .

A partir de la ecuación diferencial para la distribución de probabilidad bajo la suposición de tiempos largos (ec. A.41), planteamos condiciones de contorno sobre el problema de que la probabilidad fuera de la región de confinamiento sea nula $P(X \notin S) = 0$.

Para este caso el código de la caminata ya necesita una interacción ya que es necesario comprobar la posición de la partícula luego de que de cada paso.

Cuando suponemos que la ecuación de difusión es válida solo a tiempos largos, estamos considerando que la distribución de cada paso es discreta y que estamos esperando suficiente tiempo entre mediciones como para que la partícula de muchos pasos. Pero si suponemos inicialmente que la distribución de los pasos de una partícula es continua, e incluso podemos suponer que es normal, entonces es posible utilizar la distribución normal como la distribución para la distancia al cuadrado.

En estos casos la varianza de las caminatas va como:

$$\text{var}(\mathbf{X}) = R_0^2 \left(1 - A_1 \exp\left[\frac{-2dA_2 t}{R_0^2}\right] \right) \quad (\text{A.58})$$

donde R_0^2 es el radio de giro de la región S en la que se mueve la partícula y A_i son parámetros que caracterizan la geometría.

A.2.5. Caminatas dirigidas

[Othmer et al., 1988] Hasta ahora consideramos los casos en los que el término de correlación entre los pasos en la A.24 eran nulos, veamos que ocurre en el caso contrario.

$$c_{i,j} = \text{cov}(X_i, X_j), \quad (\text{A.59})$$

y asumiendo que el sistema está en un estado estacionario de manera que hay invariancia en traslaciones temporales $c_{i,j} = c(|i-j|)$. Renombrando $i-j=m$, podemos reescribir la ecuación A.24 como

$$\text{var}(S_n) = n \text{var}(X) + 2 \sum_{m=1}^n c(m)(n-m), \quad (\text{A.60})$$

Consideremos ahora el caso particular en la que los pasos tienen una correlación que decrece exponencialmente con el tiempo (número de pasos), renombrando $\text{var}(X) = \sigma^2$:

$$\text{cov}(X_n, X_{n+1}) = \rho \sigma^2 \quad (-1 < \rho < 1). \quad (\text{A.61})$$

es una consecuencia de la desigualdad de Schwarz A.18 que los valores de ρ vayan de -1 a 1. Entonces,

$$c(n) = \sigma^2 \rho^n$$

podemos definir un tiempo de decorrelación $n_c = -2/\log(\rho)$ de manera que se evidente que la correlación decrece exponencialmente con el número de pasos (tiempo)

$$\begin{aligned}
var(S_n) &= n \sigma^2 + 2 \sigma^2 \sum_{m=1}^{n-1} \rho^m (n-m) \\
&= \sigma^2 \left(-n + 2 \sum_{m=0}^{n-1} \rho^m (n-m) \right) \\
&= \sigma^2 \left(-n + 2 n \sum_{m=0}^{n-1} \rho^m - 2 \sum_{m=0}^{n-1} m \rho^m \right)
\end{aligned} \tag{A.62}$$

La primera suma es una serie geométrica

$$Z_n = \sum_{m=0}^{n-1} \rho^m = \frac{1-\rho^n}{1-\rho}$$

la segunda se puede evaluarse derivando la ecuación anterior respecto de ρ

$$\frac{dZ_n}{d\rho} = \sum_{m=0}^{n-1} m \rho^{m-1} = \frac{1}{\rho} \sum_{m=0}^{n-1} m \rho^m$$

por lo tanto

$$\begin{aligned}
\sum_{m=0}^{n-1} m \rho^m &= \rho \frac{dZ_n}{d\rho} = \rho \frac{d}{d\rho} \left(\frac{1-\rho^n}{1-\rho} \right) \\
&= \rho \frac{(1-\rho)(-n \rho^{n-1}) + (1-\rho^n)}{(1-\rho)^2}
\end{aligned}$$

Reemplazando en A.62:

$$\begin{aligned}
\frac{var(S_n)}{\sigma^2} &= -n + 2 n \frac{1-\rho^n}{1-\rho} + 2 \frac{n \rho^n}{1-\rho} - 2 \frac{\rho (1-\rho^n)}{(1-\rho)^2} \\
&= n \frac{1+\rho}{1-\rho} - 2 \frac{\rho (1-\rho^n)}{(1-\rho)^2},
\end{aligned} \tag{A.63}$$

$$var(S_n) = n \sigma^2 \left(\frac{1+\rho}{1-\rho} - \frac{2}{n} \frac{\rho (1-\rho^n)}{(1-\rho)^2} \right), \tag{A.64}$$

o, reescribiéndola llegamos a:

El primer término es el término asociado a la difusión y el segundo término es el término de advección o término balístico. Este segundo término tiende a 0 cuando $n \rightarrow \infty$, y la difusión tiende de manera asintótica a:

$$var(S_n) = 2 D \tau n, \tag{A.65}$$

donde definimos el coeficiente de difusión D como $\left(D_0 = \frac{\sigma^2}{2\tau}\right)$

$$\frac{D}{D_0} = \frac{1+\rho}{1-\rho}. \quad (\text{A.66})$$

Para caminatas de pocos pasos el término de advección domina, pero eventualmente termina ganando el término difusivo.

Es fácil ver que el coeficiente de difusión D crece monótonamente con ρ . $D = 0$ cuando $\rho = -1$ y $D \rightarrow \infty$ cuando $\rho \rightarrow 1$. A continuación vamos a ver qué ocurre cuando la correlación $\rho \sim \epsilon$. Consideremos $\rho = 1 - \epsilon$, $\epsilon \rightarrow 0$. Entonces

$$n_c = -\frac{2}{\log \rho} \sim \frac{2}{\epsilon},$$

$$\frac{1+\rho}{1-\rho} \sim \frac{2}{\epsilon} = n_c, \quad \frac{2}{n} \frac{\rho(1-\rho^n)}{(1-\rho)^2} \sim -\frac{n_c^2}{2n} \left(\exp\left[-\frac{2n}{n_c}\right] - 1 \right),$$

Llegando a que A.64 toma la forma

$$\text{var}(S_n) = (\sigma n_c)^2 \left[\frac{n}{n_c} + \frac{1}{2} \left(\exp\left[-\frac{2n}{n_c}\right] - 1 \right) \right], \quad \rho \rightarrow 1. \quad (\text{A.67})$$

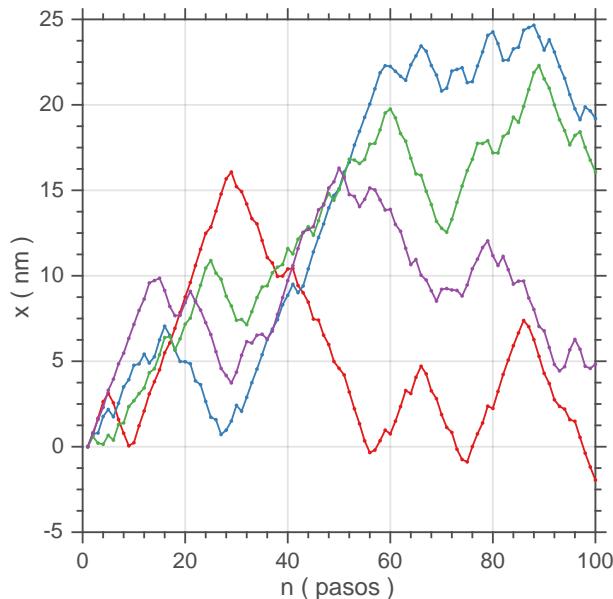


Figura A.6: Caminatas unidimensionales en las cuales los pasos están correlacionados.

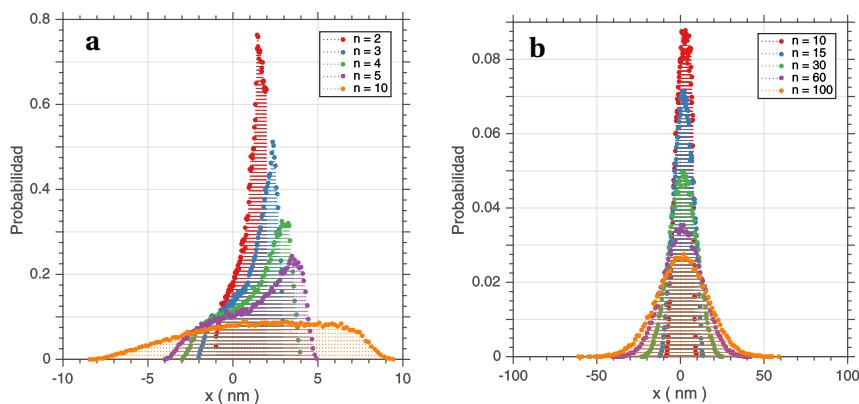


Figura A.7: (a) y (b) Histogramas de la probabilidad de encontrar al caminante luego de n pasos. Se ve como a tiempos cortos la probabilidad no es simétrica pero a medida se dan más pasos toman la forma habitual de una campana. (c) Valor medio de la posición de los caminantes y (d) la varianza, en función del número de pasos . Durante los primeros pasos el movimiento tiene una tendencia a avanzar y llega un punto en el que la correlación decae totalmente y el comportamiento es Browniano.

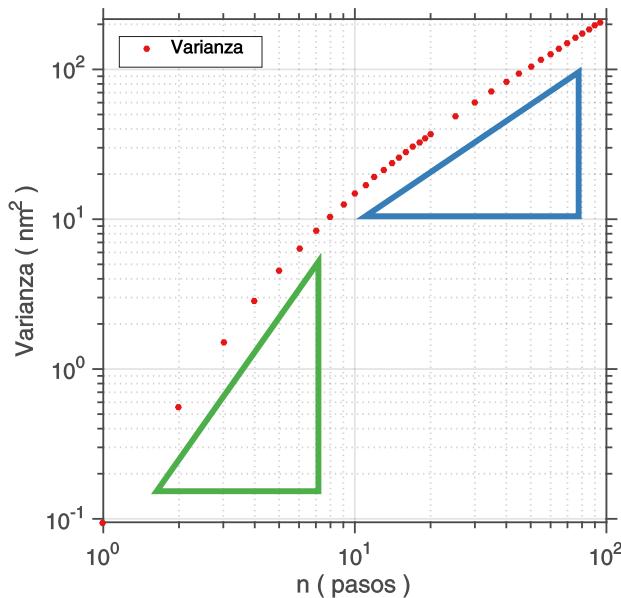


Figura A.8: Varianza en función del número de pasos. Esta imagen es la misma que A.7 (c), pero en escala logarítmica, de forma de que son evidentes los dos comportamientos. En el primer tramo, hay una pendiente de equivalente al movimiento dirigido y luego hay un cambio a una pendiente característica del movimiento Browniano.

Una de las suposiciones de las caminatas aleatorias era que cada paso era independiente de los demás, pero podemos relajar esta condición para obtener un nuevo tipo de caminatas. Esto es mucho más fácil de visualizar en dos dimensiones por lo que vamos a desarrollarlo en ese espacio.

Podemos suponer que las caminatas pueden dar pasos aleatorios pero ahora con la condición de que el ángulo de giro respecto al paso anterior esté acotado en un rango. En un experimento lo podemos observar si

tenemos suficiente resolución espacial y temporal de manera que al seguir la partícula vemos qué las mediciones que obtenemos están correlacionadas entre ellas ya que el movimiento de la partícula no cambia tan bruscamente ya que lleva cierta inercia.

Para analizar estos casos es más fácil considerar el problema a partir de las ecuaciones diferenciales.

Pero la distribución de probabilidad que describe las caminatas en las cuales hay correlación de los pasos, es la ecuación del telégrafo.

La solución para el valor medio de esta ecuación, sigue siendo nula y la varianza viene dada por:
varianza telégrafo.

Esta ecuación define un parámetro lambda que se relaciona con el tiempo medio con el que desaparece la correlación

Supongamos que el ángulo de giro de la partícula está acotado entre $[-\pi/2, \pi/2]$, esto hace que la partícula tenga una especie de preferencia para moverse en la misma dirección. Pero esto no restringe a que si observamos la distribución angular de los pasos ésta no sea uniforme en todo el intervalo $[0, 2\pi)$, ya que si la partícula se mueve por suficiente tiempo es capaz de moverse en cualquier dirección.

Para analizar este caso bidimensional útil definir las siguientes variables aleatorias. El ángulo de giro respecto de la m -ésima posición ϕ_m y la longitud del paso l_m . Podemos suponer que estas variables son independientes, de forma que tenemos:

$$\langle l \rangle = \int_0^\infty l p(l) d l, \quad (\text{A.68})$$

$$\langle l^2 \rangle = \int_0^\infty l^2 p(l) d l \quad (\text{A.69})$$

$$c = \langle \cos(\phi) \rangle = \int_{-\pi}^{\pi} \cos(\phi) g(\phi) d \phi, \quad (\text{A.70})$$

$$s = \langle \sin(\phi) \rangle = \int_{-\pi}^{\pi} \sin(\phi) g(\phi) d \phi \quad (\text{A.71})$$

donde $g(\phi)$ es la distribución de los ángulos de giro.

$$\langle S_n^2 \rangle = n \langle l^2 \rangle + 2 \langle l \rangle^2 \left(\frac{n(c - c^2 - s^2) - c}{(1 - c)^2 + s^2} + \frac{2s^2 + (c^2 + s^2)^{(n+1)/2} \gamma}{((1 - c)^2 + s^2)^2} \right), \quad (\text{A.72})$$

$$\gamma = ((1 - c)^2 - s^2) \cos((n + 1)\varphi_0) - 2s(1 - c) \sin((n + 1)\varphi_0)$$

donde $g(\phi)$ es la distribución angular de los pasos en el caso que la distribución de ángulos de giro sea simétrica $s = 0$, de manera que en estos casos la varianza está dada por

$$\text{var}(S_n) = \langle l \rangle^2 \left(n \left(\frac{1 + c}{1 - c} + b^2 \right) - \frac{2c(1 - c^n)}{(1 - c)^2} \right). \quad (\text{A.73})$$

Como primer caso, consideremos el movimiento en \mathbb{Z}^2 , de cada paso tiene longitud ℓ , y ahora, la distribución de ángulos de giro es $p(\phi = k\pi/2) = 1/4$ para $k = 0, 1, 2$ o 3 . Evaluando las fórmulas de los valores de expectación, tenemos que

$$\langle l \rangle = \ell, \quad \langle l^2 \rangle = \ell^2 \quad (\text{A.74})$$

$$c = \langle \cos(\phi) \rangle = 1/4(\cos(0) + \cos(\pi/2) + \cos(\pi) + \cos(3\pi/2)) = 0, \quad (\text{A.75})$$

$$s = \langle \sin(\phi) \rangle = \dots = 0 \quad (\text{A.76})$$

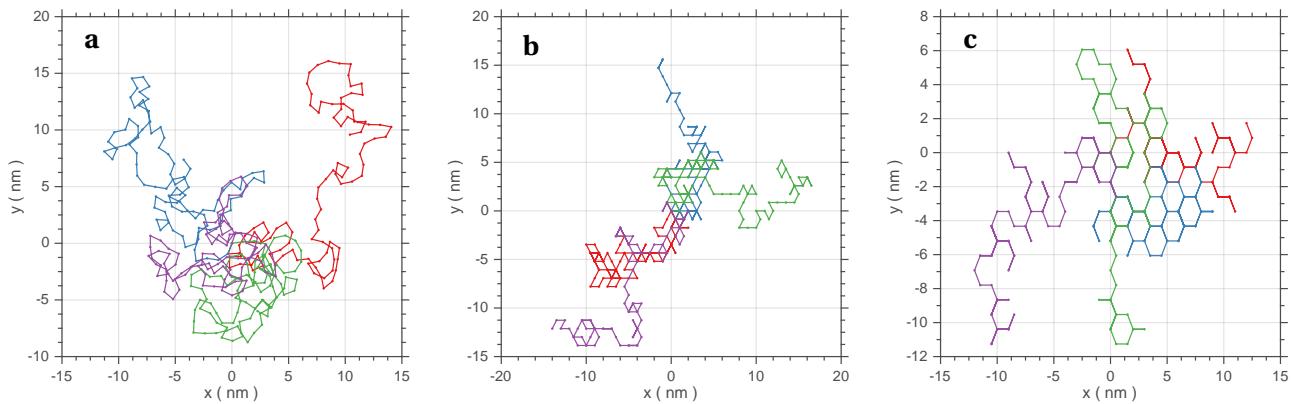


Figura A.9: Caminatas de pasos correlacionados, en dos dimensiones. (a) Correlación generada al imponer que los pasos tengan un giro máximo de 90° . (b) Los pasos solo pueden dar giros en múltiplos de 60° que da como resultado un *lattice* triangular. (c) Los pasos solo pueden dar giros en múltiplos impares de 60° que da como resultado un *lattice* hexagonal.

Si ahora podemos considerar el caso en el que el *lattice* sobre el que se mueve es una grilla triangular de lado L , en este caso la distribución de probabilidad del ángulo de giro es $p(\phi = k\pi/6) = 1/6$ con $k = 0, 1, \dots, 5$.

Un último ⁴ ejemplo consideramos el *lattice* hexagonal. Para este caso la distribución de probabilidad es $p(\phi = k\pi/3)$ con $k = 1, 3$ y 5 . Saliendo de los casos de los *lattice*, podemos considerar el caso de tener una distribución uniforme pero que no completa una vuelta entera, es decir $[-\alpha, \alpha]$. Este caso $c = \sin(\alpha)/\alpha$

A.2.6. Ecuación del telégrafo

$$\begin{aligned} \frac{\partial p^+}{\partial t} + s \frac{\partial p^+}{\partial x} &= -\lambda p^+ + \lambda p^- \\ \frac{\partial p^-}{\partial t} - s \frac{\partial p^-}{\partial x} &= \lambda p^+ - \lambda p^- \end{aligned} \quad (\text{A.77})$$

⁴Estos son los tres únicos casos posibles de tesselaciones con polígonos regulares.

$$\begin{aligned}\frac{\partial p}{\partial t} + \frac{\partial j}{\partial x} &= 0 \\ \frac{\partial j}{\partial t} + 2\lambda j &= -s^2 \frac{\partial p}{\partial x}\end{aligned}\tag{A.78}$$

$$\frac{\partial^2 p}{\partial t^2} + 2\lambda \frac{\partial p}{\partial t} = s^2 \frac{\partial^2 p}{\partial x^2}\tag{A.79}$$

$$\frac{d^2 \langle x^2 \rangle}{dt^2} + 2\lambda \frac{d \langle x^2 \rangle}{dt} = 2s^2\tag{A.80}$$

$$\langle x^2(t) \rangle = \frac{s^2}{\lambda} \left\{ t - \frac{1}{2\lambda} (1 - \exp[-2\lambda t]) \right\}\tag{A.81}$$

$$D = \frac{s^2}{2\lambda} = \frac{\lambda \Delta^2}{2}\tag{A.82}$$

$$p(x, t) = \begin{cases} \frac{\exp[-\lambda t]}{2} \{ \delta(x - st) + \delta(x + st) + \frac{\lambda}{s} [I_0(\Lambda) + \frac{\lambda t}{\Lambda} I_1(\Lambda)] \} & \text{para } |x| < st \\ 0 & \text{para } |x| \geq st \end{cases}\tag{A.83}$$

$$\Lambda = \lambda \sqrt{t^2 - x^2/s^2}$$

$$\begin{aligned}I_0 &\sim \frac{e^x}{\sqrt{2\pi x}} + \mathcal{O}(1/x) \\ I_1 &\sim \frac{e^x}{\sqrt{2\pi x}} + \mathcal{O}(1/x)\end{aligned}\tag{A.84}$$

$$p(x, t) \sim \frac{1}{\sqrt{4\pi D t}} \exp\left[\frac{-x^2}{4Dt}\right] + \exp[-\lambda t] \mathcal{O}(\xi^2)\tag{A.85}$$

$$\xi \equiv x^2/(st)^2$$

$$p(x, t) \equiv p^+(x, t) + p^-(x, t)$$

$$j \equiv s(p^+ - p^-) p(x, 0) = p_0(x), \quad j(x, 0) = j_0(x) \quad p(x, 0) = p_0(x), \quad \partial p / \partial t(x, 0) = -\partial j_0 / \partial x(x)$$

A.2.7. Distribución de probabilidad para caminatas cortas

Una pregunta que surgió durante el trabajo es a partir de qué valor es válida la aproximación de Rayleigh para la probabilidad, [Rayleigh, 1919, Hughes, 1996]

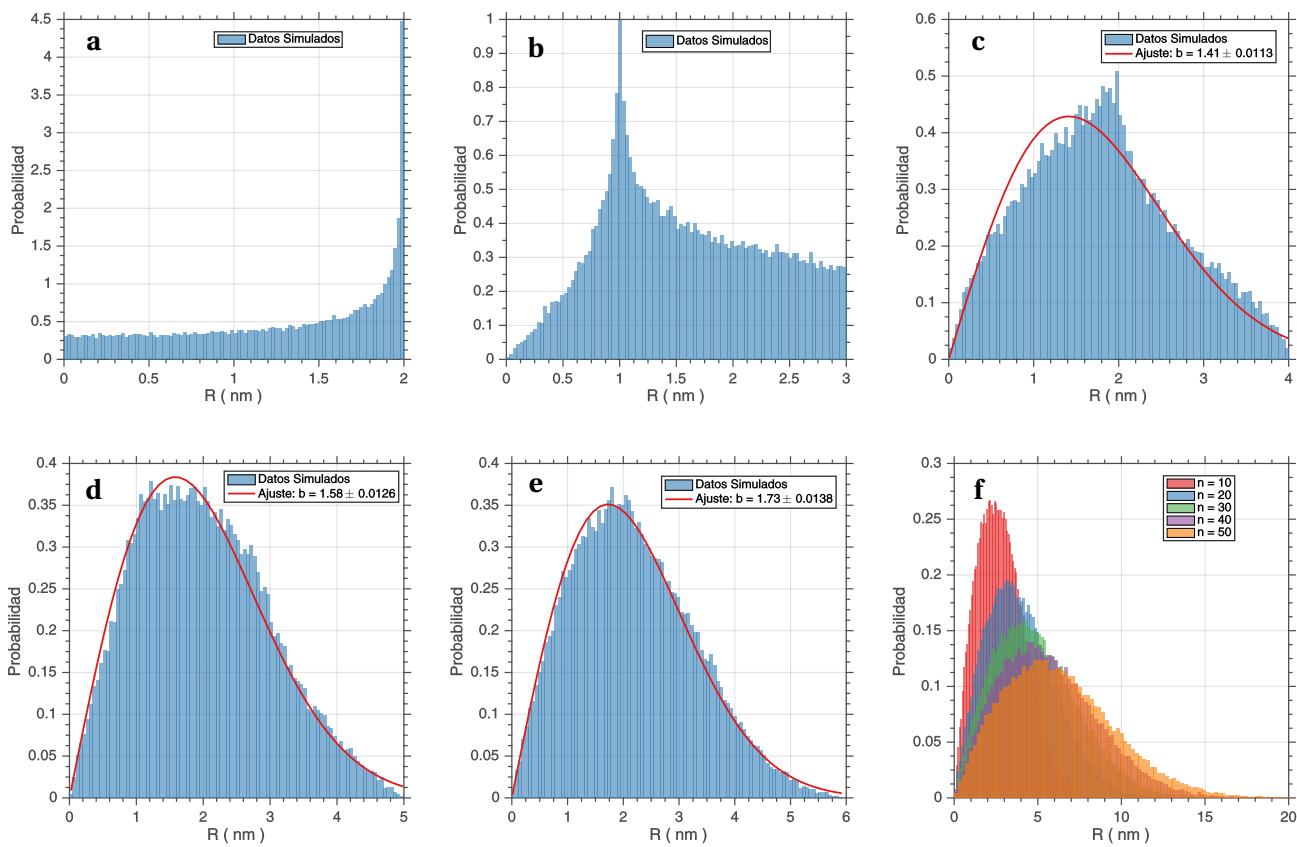


Figura A.10: Densidades de probabilidad para caminatas bidimensionales de paso fijo $\ell = 1$ (*Pearson's walk*). En (c - e) se grafican los ajustes de la distribución de Rayleigh para iniciar las diferencias con las mismas. A partir de 7 pasos, los datos simulados son aceptados (mediante un Lilliefors test, $\alpha = 0,05$) como pertenecientes a una distribución de Rayleigh con el parámetro estimado de los datos.

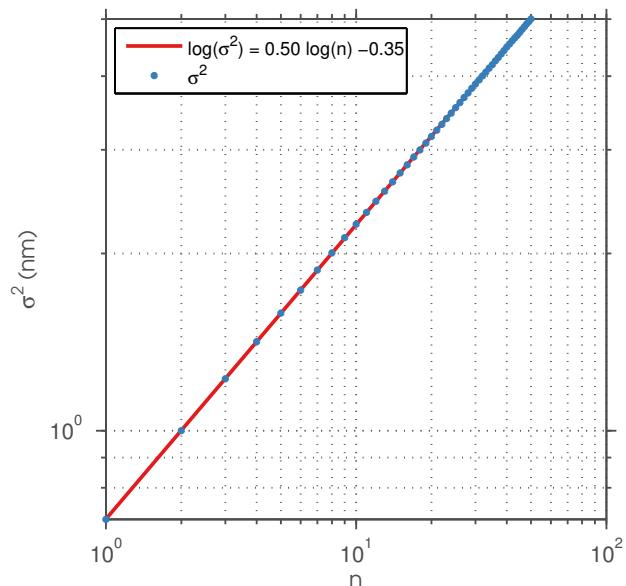


Figura A.11: Valores del parámetro de ajuste en función del número de pasos, junto con el ajuste.

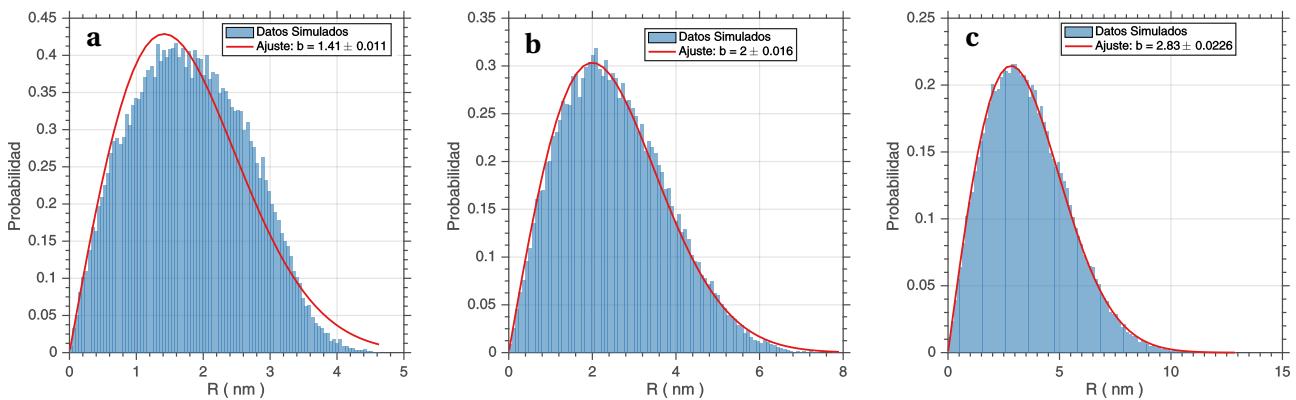


Figura A.12: Histogramas de las probabilidades de encontrar al caminante a una distancia R luego de (a) 2 pasos , (b) 4 pasos, (c) 8 pasos. Las caminatas tiene distribución de pasos uniformes en cada dirección. Al igual que en el caso de las caminatas de longitud de paso fijo, la aproximación de Rayleigh es aceptada (en un test de Lilliefors, $\alpha = 0,05$) a partir de los 7 pasos.

$$p(\mathbf{r}) = \frac{1}{2\pi|\mathbf{r}|} \omega(|\mathbf{r}|). \quad (\text{A.86})$$

A.2.8. Caminatas de paso fijo (Pearson Walks)

$$p_j(\mathbf{r}) = \frac{1}{2\pi|\mathbf{r}|} \delta(|\mathbf{r}| - a_j). \quad (\text{A.87})$$

$$P_n(\mathbf{r}) = \psi_n(|\mathbf{r}|) = \frac{1}{2\pi} \int_0^\infty z J_0(z|r|) \prod_{i=1}^n J_0(za_i) dz. \quad (\text{A.88})$$

$$\Pr\{r \leq R, n\} = 2\pi \int_0^R \rho \psi_n(\rho) d\rho \quad (\text{A.89})$$

$$= R \int_0^\infty z J_1(zR) \prod_{i=1}^n J_0(za_i) dz \quad (n \geq 2). \quad (\text{A.90})$$

Hay un resultado exacto muy sencillo, obtenido por Kluyver [Kluyver, 1906], que es válido en el caso de que todos los pasos tengan longitud a .

$$\Pr\{a, n\} = \frac{1}{n+1}. \quad (\text{A.91})$$

$$P_n(\mathbf{r}) = \psi_n(|\mathbf{r}|) \sim \frac{1}{\pi n a^2} \exp\left[-\frac{r^2}{na^2}\right], \quad (\text{A.92})$$

donde

$$a^2 = \frac{1}{n} \sum_{i=1}^n a_i^2, \quad (\text{A.93})$$

A.2.9. Caminatas de paso variable

En el caso que los pasos están idénticamente distribuidos de manera isotrópica en el espacio, siguiendo la ec. A.86

$$\chi(|q|) = \int_0^\infty J_0(|q|r) \omega(r) dr. \quad (\text{A.94})$$

$$P_n(\mathbf{r}) = \psi_n(|\mathbf{r}|) = \frac{1}{2\pi} \int_0^\infty z J_0(z|r|) \chi(z)^n dz. \quad (\text{A.95})$$

$$\Pr\{r \leq R, n\} = R \int_0^\infty z J_1(zR) \chi(z)^n dz. \quad (\text{A.96})$$

Un caso particular de interés es en el que los pasos están distribuidos de manera Gaussiana,

$$p(\mathbf{r}) = \frac{1}{\pi a^2} \exp\left[-\frac{|\mathbf{r}|^2}{a^2}\right] \quad (\text{A.97})$$

$$\omega(r) = \frac{2r}{\pi a^2} \exp\left[-\frac{r^2}{a^2}\right]. \quad (\text{A.98})$$

$$\psi_n(r) = \frac{1}{\pi n a^2} \exp\left[-\frac{r^2}{n a^2}\right]. \quad (\text{A.99})$$

Los resultados anteriores se pueden generalizar bajo la suposición de que la varianza de la longitud de los pasos es finita

$$\sigma^2 = \text{var}(X) = \int_0^\infty \ell^2 \omega(\ell) d\ell. \quad (\text{A.100})$$

Y en el caso de muchos pasos, obtenemos que la distribución se aproxima por

$$P_n(\mathbf{r}) = \psi_n(|\mathbf{r}|) = \frac{1}{\pi n \sigma^2} \exp\left[-\frac{r^2}{n \sigma^2}\right]. \quad (\text{A.101})$$

$$\Pr\{r \leq R, n\} \sim \frac{R^2}{n \sigma^2} \quad \text{cuando } n \rightarrow \infty. \quad (\text{A.102})$$

Para los resultados ver para caminatas en tres dimensiones ir a ([Hughes, 1996] 2.5).

B

Códigos computacionales

A modo de ejemplo presentamos el código completo de una función que genera caminatas con pasos distribuidos de manera gaussiana. Todas las demás funciones y aplicaciones se encuentran en el repositorio de Github `randomWalk-matlab`[Kielbowicz, 2017]

```
1 function out = randomWalkGaussian(nStep,meanStep,stdStep,varargin)
2 % RANDOMWALKGAUSSIAN Generador de caminatas de pasos con distribucion
3 % Gaussiana.
4
5 %W = RANDOMWALKGAUSSIAN(N) caminata de N pasos en dos dimensiones. Con
6 % pasos con distribucion Norm(0,1)
7 %W = RANDOMWALKGAUSSIAN(N,MEAN,STD) caminata de N pasos en dos dimensiones. Con
8 % pasos con distribucion Norm(MEAN,STD) en dos dimensiones
9 %W = RANDOMWALKGAUSSIAN(N,MEAN,STD,'Dimension',DIM) caminata de N pasos en dos dimensiones. Con
10 % pasos con distribucion Norm(MEAN,STD) en DIM dimensiones
11 % Todas las caminatas parten del origen (0,...,0), por lo que W tiene longitud N+1
12 % donde N es el numero de pasos
13
14 %Defaults
15 nDimension = 2;
16 if nargin < 2
17 meanStep = 0;
18 stdStep = 1;
19 end
20
21 if numel(varargin)>0
22
23 for i=1:numel(varargin)
24 if strcmpi(varargin{i}, 'Dimension')
25 nDimension = varargin{i+1};
26 break
27 end
28 end
29
30 end
31
32 randomStep = meanStep(ones(nStep,nDimension)) + stdStep randn(nStep,nDimension);
33 out = [zeros(1,nDimension);cumsum(randomStep)];
```

Listing B.1: Generador de caminatas gaussianas

A continuación damos una lista de las funciones definidas en la caja de herramientas de `randomWalk-matlab`.

```
1 randomWalkGaussian % genera caminatas con pasos gaussianos
2 randomWalkConfined % genera caminatas confinadas
3 randomWalkCorrelated % genera caminatas con correlacion/dirigidas en funcion del \angulo de
4 giro
randomWalkDistribution % genera caminatas con pasos con la distribucion dada
```

B. Códigos computacionales

```
5 randomWalkFixStep % genera caminatas con pasos de longitud constante
6 randomWalk % constructor de la clase randomWalk
7 importJournal % importa la informaci\'on del .jrn de SymFCS
8 importWalk % importa las posiciones de la trayectoria a partir del .txt generado por SymFCS
9 msd % calcula el msd de una caminata
10 theoreticalMSD % devuelve la funci\'on del msd dado para evaluar
11 segmentCluster % segmentaci\'on
12 segmentThread % segmentacion de la caminata a partir del angulo de giro entre pasos consecutivos
13 segmentByAnisotropy % segmentaci\'on de la caminata a utilizando la anisotrop\'ia local
14 guiSegment % interfaz gr\'afica para ver los segmentos de una caminata
15 plotTheoreticalMSD % grafica la curva de msd te\lorica
16 plotAngular % grafica los histogramas polares y da un resumen de los par\'ametros angulares de la
   caminata
17 plotScatter % gr\'afico de dispersi\'on
18 plotGMMID % grafica el histograma de los datos junto con las curvas estimadas por el GMM
19 plotGMM2D % gr\'afico de dispersi\'on de los datos junto con las curvas de nivel de las
   distribuciones estimadas por GMM
20 gmmNComponentLikelihood % eval\'ua y grafica el AIC y BIC con distinta cantidad de componentes
21 plotMovie % grafica la animaci\'on caminata, tiene opci\'on para pausar
22 saveVideo % guarda la animaci\'on de la caminata en formato mp4
23 formatFigure %Da formato a las figuras
```

B.1. randomWalk()

B.1.1. Propiedades

```
1 positions % Posiciones importadas del .txt
2 journal % informaci\'on extra\'ida del .jrn
3 oldPositions % respaldo de las posiciones para cuando se eliminan puntos o tramos de la
   trayectoria
4 name % nombre del archivo de la medici\'on
5 segment % objeto donde se guardan los \\'indices y tama\~nos de los segmentos.
6 pixelSize % tama\~no de pixel
7 stepTime % tiempo entre mediciones
8 timeUnit % unidad temporal
9 spaceUnit % unidad espacial
10 voltageZoom % tabla con voltajes aplicados a los espejos, se relaciona con pixelScale
11 pixelScale % tabla con los tama\~nos de pixel para los distintos voltajes. !!Se obtienen a partir
   de la calibraci\'on.
```

B.1.2. M\'etodos disponible en la clase

Aqu\' presentamos una lista de todos los m\'etodos disponibles en la clase, para la informaci\'on completa del uso de los m\'etodos se puede ver la documentaci\'on de las mismas usando `help` o `doc` de Matlab Los m\'etodos pueden aplicarse como una funci\'on `anisotropy(walk)` o con la notaci\'on de punto `walk.anisotropy()`.

```
1
2 %% Constructor
3 randomWalk % constructor de la clase
4 subsref(obj, s) % para referenciar a porciones de la caminata a partir del \'indice usando la
   misma notaci\'on de dos puntos (':') que para vectores walk(10:end)
5
6 %%M\'etodos set para las propiedades
7
8 setName % asigna el nombre
9 setPixelSize % asigna el tama\~no de pixel
```

```

10 setStepTime % asigna el tiempo entre mediciones
11 setUnits % asigna las unidades
12
13 %%M\ 'etodos para limpiar la caminata
14 removePoints(obj, pointsIndex) % elimina puntos de la caminata dados su indice
15 removeZeroLengthStep(obj, varargin) % elimina todos las posiciones de la caminata donde los pasos
   son nulos
16 restoreOldPositions % reestablece las posiciones al estado anterior de utilizar removePoints o
   removeZeroLengthStep
17 restoreOriginalPositions % reestablece las posiciones al estado adquirido al construir o
   inicializar el objeto
18
19 %%M\ 'etodos de adquisici\on
20
21 t % vector con los tiempos de cada medici\on
22 x % vector con las posiciones en la direcci\on x
23 y % vector con las posiciones en la direcci\on y
24 z % vector con las posiciones en la direcci\on z
25 XY % matriz de nPasos x 2 con las posiciones en el plano XY
26 dimension % dimension en la que ocurre la caminata
27
28 %%M\ 'etodos de c\alculo
29
30 D % coeficiente de difusión
31 anisotropy % anisotrop\ia de la caminata
32 duration(obj, unit) % duraci\on de la caminata en unidades especificadas por 'unit'
33 meanCosine % valor medio del coseno de \angulo de giro
34 meanStep % valor medio de los pasos
35 msd(obj, fraction) % desplazamiento cuadr\atrico medio de la caminata, la funci\on devuelve los
   valores hasta nPasos fraction
36 radius % extension de la caminata medida desde el centro de momento/masa de la misma hasta el
   punto m\'as alejado a este
37 stepAngle % \angulos direccionales de los pasos
38 stepLength % longitudes de los pasos
39 stepVariance % varianza de los pasos
40 steps % devuelve los pasos
41 stepsNumber % n\'umero de pasos de la caminata
42 v % valor de la velocidad de deriva
43
44 %%M\ 'etodos de graficos
45
46 plot % grafico de la caminata
47 plotAngular % grafico y resumen de los \angulos direccionales y de giro
48 plotScatter % grafico de dispersi\on polar de los pasos junto con histogramas de las proyecciones
   de los pasos en los ejes x e y
49 plotTheoreticalMSD(obj, type, fraction) % grafico de la curva de msd te\orica utilizando los
   parametros estimados por los otros m\etodos de la clase
50
51 %%M\ 'etodos de segmentaci\on
52
53 segmentByAnisotropy % segmentacion a partir de la anisotropia
54 segmentCluster % segmentacion a partir de la probabilidad de escape para adquirir los c\'umulos
55 segmentThread % segmentacion a partir del \angulo de giro para adquirir los filamentos

```


Bibliografía

- [Bachelier, 1900] Bachelier, L. (1900). The theory of speculation. *Annales scientifiques de l'École Normale Supérieure*, 3(17):21–86.
- [Behboodian, 1970] Behboodian, J. (1970). On the Modes of a Mixture of Two Normal Distributions. *Technometrics*, 12(1):131.
- [Berens, 2009] Berens, P. (2009). CircStat : A MATLAB Toolbox for Circular Statistics. *Journal of Statistical Software*, 31(10).
- [Berens and Velasco, 2009] Berens, P. and Velasco, M. J. (2009). Circstat: Circular Statistics Toolbox. <https://github.com/circstat>.
- [Bevington and Robinson, 1992] Bevington, P. R. and Robinson, D. K. (1992). *Data Reduction and Error Analysis for the Physical Sciences*. McGraw-Hill Education.
- [Boyer et al., 2013] Boyer, D., Dean, D. S., Mejía-Monasterio, C., and Oshanin, G. (2013). Optimal least-squares estimators of the diffusion constant from a single Brownian trajectory. *European Physical Journal: Special Topics*, 216(1):57–71.
- [Bressloff, 2014] Bressloff, P. C. (2014). *Stochastic Processes in Cell Biology*, volume 41 of *Interdisciplinary Applied Mathematics*. Springer International Publishing, Cham.
- [Brown, 1829] Brown, R. (1829). XXIV. Additional Remarks On Active Molecules. *Philosophical Magazine Series 2*, 6(33):161–166.
- [Chandrasekhar, 1943] Chandrasekhar, S. (1943). Stochastic problems in physics and astronomy.
- [Codling and Hill, 2005] Codling, E. and Hill, N. (2005). Sampling rate effects on measurements of correlated and biased random walks. *Journal of Theoretical Biology*, 233(4):573–588.
- [Codling et al., 2008] Codling, E. A., Plank, M. J., and Benhamou, S. (2008). Random walk models in biology. *Journal of The Royal Society Interface*, 5(25):813–834.
- [Coppin et al., 1996] Coppin, C. M., Finer, J. T., Spudich, J. a., and Vale, R. D. (1996). Detection of sub-8-nm movements of kinesin by high-resolution optical-trap microscopy. *Proceedings of the National Academy of Sciences*, 93(5):1913–1917.
- [Coscoy et al., 2007] Coscoy, S., Huguet, E., and Amblard, F. (2007). Statistical Analysis of Sets of Random Walks: How to Resolve Their Generating Mechanism. *Bulletin of Mathematical Biology*, 69(8):2467–2492.
- [Durr et al., 2007] Durr, N. J., Larson, T., Smith, D. K., Korgel, B. A., Sokolov, K., and Ben-Yakar, A. (2007). Two-Photon Luminescence Imaging of Cancer Cells Using Molecularly Targeted Gold Nanorods. *Nano Letters*, 7(4):941–945.

- [Einstein, 1905] Einstein, A. (1905). Über die von der molekularkinetischen Theorie der Wärme geforderte Bewegung von in ruhenden Flüssigkeiten suspendierten Teilchen. *Annalen der Physik*, 322(8):549–560.
- [Feller, 1968] Feller, W. (1968). An Introduction to Probability Theory and Its Applications.
- [Gabriel, 2015] Gabriel, M. (2015). *Caracterización de nanopartículas metálicas como sondas alternativas para microscopía de fluorescencia y estudio de estructura en la nanoescala*. PhD thesis, Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires.
- [Gnedenko and Kolmogorov, 1954] Gnedenko, B. V. and Kolmogorov, A. N. (1954). *Limit distributions for sums of independent random variables*. Addison-Wesley Mathematics Series. Addison-Wesley, Cambridge, MA.
- [Gordon, 1972] Gordon, M. (1972). Modern theory of polymer solutions. H. Yamakawa. Harper & Row, London. 1971. pp. 419. Price £9.35. *British Polymer Journal*, 4(6):541–542.
- [Hand et al., 1989] Hand, D. J., McLachlan, G. J., and Basford, K. E. (1989). Mixture Models: Inference and Applications to Clustering. *Applied Statistics*, 38(2):384.
- [Hogg et al., 2014] Hogg, R. V., Tanis, E., and Zimmerman, D. (2014). *Probability and Statistical Inference*. Pearson.
- [Hughes, 1996] Hughes, B. (1996). Random walks and random environments.
- [Kac, 1947] Kac, M. (1947). Random Walk and the Theory of Brownian Motion. *The American Mathematical Monthly*, 54(7):369.
- [Kareiva and Shigesada, 1983] Kareiva, P. M. and Shigesada, N. (1983). Analyzing insect movement as a correlated random walk. *Oecologia*, 56(2-3):234–238.
- [Kielbowicz, 2017] Kielbowicz, A. (2017). Random Walk Analysis Toolbox. <https://github.com/akielbowicz/randomWalk-matlab>.
- [Kluyver, 1906] Kluyver, J. (1906). A local probability problem. *Nederl. Acad. Wetensch. Proc*, 8:341—350.
- [Kusumi et al., 1993] Kusumi, A., Sako, Y., and Yamamoto, M. (1993). Confined lateral diffusion of membrane receptors as studied by single particle tracking (nanovid microscopy). Effects of calcium-induced differentiation in cultured epithelial cells. *Biophysical Journal*, 65(5):2021–2040.
- [Kusumi et al., 2014] Kusumi, A., Tsunoyama, T. A., Hirosawa, K. M., Kasai, R. S., and Fujiwara, T. K. (2014). Tracking single molecules at work in living cells. *Nature Chemical Biology*, 10(7):524–532.
- [Landau and Lifshitz, 1969] Landau, L. D. and Lifshitz, E. M. (1969). *Statistical Physics, Third Edition, Part 1: Volume 5 (Course of Theoretical Physics, Volume 5)*. Butterworth-Heinemann; 3 edition (January 15, 1980).
- [Levi and Gratton, 2007] Levi, V. and Gratton, E. (2007). Exploring dynamics in living cells by tracking single particles. *Cell Biochemistry and Biophysics*, 48(1):1–15.
- [Martin et al., 2002] Martin, D. S., Forstner, M. B., and Käs, J. a. (2002). Apparent Subdiffusion Inherent to Single Particle Tracking. *Biophysical Journal*, 83(4):2109–2117.
- [Maucort et al., 2014] Maucort, G., Kasula, R., Papadopoulos, A., Nieminen, T. A., Rubinsztein-Dunlop, H., and Meunier, F. A. (2014). Mapping organelle motion reveals a vesicular conveyor belt spatially replenishing secretory vesicles in stimulated chromaffin cells. *PLoS ONE*, 9(1):1–9.

- [Michalet, 2010] Michalet, X. (2010). Mean square displacement analysis of single-particle trajectories with localization error: Brownian motion in an isotropic medium. *Physical Review E*, 82(4):041914.
- [NIST, 2012] NIST (2012). *NIST/SEMATECH e-Handbook of Statistical Methods*. SEMATECH.
- [Othmer et al., 1988] Othmer, H. G., Dunbar, S. R., and Alt, W. (1988). Models of dispersal in biological systems. *Journal of Mathematical Biology*, 26(3):263–298.
- [Pearson, 1905] Pearson, K. (1905). The problem of the random walk. *Nature*, page 294.
- [Qian et al., 1991] Qian, H., Sheetz, M., and Elson, E. (1991). Single particle tracking. Analysis of diffusion and flow in two-dimensional systems. *Biophysical Journal*, 60(4):910–921.
- [Rayleigh, 1919] Rayleigh, L. (1919). XXXI. On the problem of random vibrations, and of random flights in one, two, or three dimensions. *Philosophical Magazine Series 6*, 37(220):321–347.
- [Roberts et al., 2007] Roberts, G. S., Wood, T. A., Frith, W. J., and Bartlett, P. (2007). Direct measurement of the effective charge in nonpolar suspensions by optical tracking of single particles. *Journal of Chemical Physics*, 126(19):194503.
- [Rudnick and Gaspari, 1986] Rudnick, J. and Gaspari, G. (1986). The asphericity of random walks. *Journal of Physics A: Mathematical and General*, 19(4):L191–L193.
- [Ruhnow et al., 2011] Ruhnow, F., Zwicker, D., and Diez, S. (2011). Tracking Single Particles and Elongated Filaments with Nanometer Precision. *Biophysical Journal*, 100(11):2820–2828.
- [Rycroft and Bazant, 2005] Rycroft, S. C. H. and Bazant, M. Z. (2005). Lecture 1 : Introduction to Random Walks and Diffusion. Technical report, Department of Mathematics, MIT.
- [Saffman and Delbrück, 1975] Saffman, P. G. and Delbrück, M. (1975). Brownian motion in biological membranes. *Proceedings of the National Academy of Sciences*, 72(8):3111–3113.
- [Saxton, 1993a] Saxton, M. (1993a). Lateral diffusion in an archipelago. Dependence on tracer size. *Biophysical Journal*, 64(4):1053–1062.
- [Saxton, 1993b] Saxton, M. (1993b). Lateral diffusion in an archipelago. Single-particle diffusion. *Biophysical Journal*, 64(6):1766–1780.
- [Saxton, 1995] Saxton, M. (1995). Single-particle tracking: effects of corrals. *Biophysical Journal*, 69(2):389–398.
- [Saxton and Jacobson, 1997] Saxton, M. J. and Jacobson, K. (1997). SINGLE-PARTICLE TRACING: Applications to Membrane Dynamics. *Annual Review of Biophysics and Biomolecular Structure*, 26(1):373–399.
- [Sciutto, 1994] Sciutto, S. J. (1994). Study of the shape of random walks. *Journal of Physics A: Mathematical and General*, 27(21):7015–7034.
- [Sciutto, 1995] Sciutto, S. J. (1995). Study of the shape of random walks. II. Inertia moment ratios and the two-dimensional asphericity. *Journal of Physics A: Mathematical and General*, 28(13):3667–3679.
- [Scott, 2002] Scott, A. (2002). *Circular Data: An Overview with Discussion of One-Sample Tests*. PhD thesis, Montana State University, Bozeman, Montana.
- [Simson et al., 1995] Simson, R., Sheets, E., and Jacobson, K. (1995). Detection of temporary lateral confinement of membrane proteins using single-particle tracking analysis. *Biophysical Journal*, 69(3):989–993.

- [Smoluchowski, 1976] Smoluchowski, R. (1976). Short linear random walks. *Journal of Physics and Chemistry of Solids*, 37(1):89–93.
- [Spiegel et al., 2012] Spiegel, M. R., Schiller, J., and Srinivasan, R. A. (2012). *Schaum's Outline of Probability and Statistics*. McGraw-Hill, New York, New York, USA, 4th edition.
- [Stephens, 1974] Stephens, M. A. (1974). EDF Statistics for Goodness of Fit and Some Comparisons. *Journal of the American Statistical Association*, 69(347):730.
- [Uhlenbeck and Ornstein, 1930] Uhlenbeck, G. E. and Ornstein, L. S. (1930). On the Theory of the Brownian Motion. *Physical Review*, 36(5):823–841.
- [von Smoluchowski, 1906] von Smoluchowski, M. (1906). Zur kinetischen Theorie der Brownschen Molekularbewegung und der Suspensionen. *Annalen der Physik*, 326(14):756–780.
- [Zarr, 2010] Zarr, J. H. (2010). *Biostatistical Analysis*. Prentice Hall, fifth edition.



Bibliografía

Agradecimientos

Primero, quiero agradecer a Laura y Graciana por sus consejos e indicaciones durante el desarrollo de esta tesis. Y a todos aquellos de quienes dedicaron su tiempo y esfuerzo para enseñarme durante el transcurso de la carrera, principalmente aquellos que más me influenciaron: Miguel, Hernán, Vera, Claudia, Adriana, César, Rafael, Fernando, Ariel, Alberto, Guillermo, Gabriel y Teresa.

A mi abu Nieve por apoyarme durante todo este tiempo principalmente dandome un lugar en su hogar. A mi tía Tere y a Max por apoyarme durante toda mi vida para que llegue a esta instancia de mi formación, también a mi tia Xime por financiar estos años que estuve en Buenos Aires. A mis tíos Blanca, Magic, Luly, Betty, Vero y Camola y a mis tíos Pato, Peter, Mario, Gonzalo y Bob por todo el apoyo durante tantos años por ser grandes ejemplos a seguir. Y a Geisha y Cari por la compañía y el cariño que me brindan.

Alex, mi mamá y Gushi mi viejo por los genes, el techo y la comida. Por haberme fomentado y permitido, durante toda la vida, a ser curioso, creativo y caprichoso.

Con quien compartí desde el principio: Agus, Joan, Eze, Franco y Sebas. A Jime y Tami por los buenos momentos que pasamos, por los paseos y las mateadas. A Maru por su compañía este último año, por las comidas, la cerveza y todo lo que eso conlleva. A todos los que fueron compañeros, con quien tomé mate y pasé horas criticando todo en vez de estudiar y con quienes estudié: Belén, Christian, Guille, Guido, Jime, Juan, Juan, Juan Cruz, Juanma, , Joel, Lean, Leo, ,Mari, Nacho, Nacho, Nachito, Noe, Noe, Rami, Rodri y Yani.

A los miembros del LEC: Andrea, Hernán, Maxi, Edu, Fabri, Fede, Juan, Martín y Nico. Y especialmente a Manu por compartirme las mediciones de las trayectorias y sus dudas sobre el análisis, ya que muchos de los resultados fueron producto de sus dudas.

A todos aquellos que conforman la UBA y mantienen la institución en pie.

Quiero dedicar este trabajo a mis hermanas Nicole y Estefy, porque no tengo nada más para ofrecerles.

Nuevamente gracias a todos, por los momentos que compartimos y el cariño que me brindaron.

