



UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO

FACULTAD DE CIENCIAS

ANÁLISIS NUMÉRICO E IMPLEMENTACIÓN
COMPUTACIONAL DE LA DIMENSIÓN FRACTAL EN
SISTEMAS DINÁMICOS EN BIOLOGÍA

REPORTE DE SERVICIO SOCIAL

QUE PARA OBTENER EL TÍTULO DE:

FÍSICO

P R E S E N T A:

JOSÉ ROBERTO BASTIDA LUNA

TUTOR:

DR. PEDRO EDUARDO MIRAMONTES VIDAL
DR.



CIUDAD DE MÉXICO, 2021

Capítulo 1

Introducción

El presente reporte tiene por objeto proporcionarle a los estudiantes del curso de Biología Matemática los conceptos generales sobre:

- Las ecuaciones diferenciales discretas y los Sistemas de Funciones Iteradas (SFI).
- El Juego del Caos: Se introduce el triángulo de Sierpinski y el juego del caos aplicado a subcadenas del cromosoma 19.
- La dimensión de Minkowski-Bouligand (también conocido como la dimensión del *conteo de cajas*).

La dimensión de Minkowski-Bouligand converge a la dimensión topológica. Por tal motivo, la primera aproximación consiste en guiar al alumno por los conjuntos que ya conoce, a saber, una recta y un cuadrado con dimensiones topológicas iguales a 1 y 2 respectivamente. Se calcula también la dimensión de Minkowski-Bouligand para:

- El triángulo de Sierpinski: con el fin de validar que coincida con la dimensión fractal teórica calculada a partir de la autosemejanza.
- El atractor de Henón: un Sistema de Funciones Iteradas (SFI) cuadrático cuya dimensión teórica es de 1.26.

Finalmente se introduce el concepto del juego del caos aplicado a subcadenas de DNA de diferentes longitudes y se calcula la dimensión de Minkowski-Bouligand a los conjuntos que dimanan de ésta con el fin de ilustrar que aún sin una autosemejanza clara, es posible calcular la dimensión de cualquier subconjunto de \mathbb{R}^2 siempre que su cardinalidad sea lo suficientemente grande para que el algoritmo converga.

Todos los métodos computacionales asociados al presente proyecto así como la guía para la ejecución e implementación están disponibles en *esta liga*. Consiste en un proyecto desarrollado en el lenguaje de programación python 3.x. Se exhorta al alumno para que visite la página del proyecto, instale la versión más reciente de python 3.x y experimente con las funciones que aquí se presentan.

Capítulo 2

Sistemas dinámicos discretos y fractales

En general, las ecuaciones diferenciales discretas se pueden representar como:

$$x_{n+1} = f(\vec{\alpha}, x_n) \quad (2.1)$$

Donde $\vec{\alpha}$ es algún vector de parámetros y x_n es el estado previo del sistema. Cada estado del sistema depende del estado inmediato anterior en el que el sistema se encontró.

Dada una ecuación diferencial discreta como la que se muestra en 2.1 y un punto inicial x_0 dentro del dominio de la ecuación, uno puede conocer la dinámica del sistema. Para esto basta componer la ecuación consigo misma al rededor del punto inicial x_0 un número lo suficientemente grande de veces para que arroje información sobre la evolución del sistema:

$$\begin{aligned} & x_0 \\ & x_1 = f(\vec{\alpha}, x_0) \\ & x_2 = f \circ f(\vec{\alpha}, x_0) \\ & \dots \\ & x_n = \underbrace{f \circ \dots \circ f}_{n \text{ veces}}(\vec{\alpha}, x_0) \end{aligned} \quad (2.2)$$

A partir de estas iteraciones, se obtiene una sucesión de puntos $\mathcal{S} = \{x_0, x_1, \dots, x_n\}$ cuyo análisis permite conocer la dinámica del sistema. La relevancia que esto tiene es que permite modelar los escenarios en que el sistema evoluciona y comprender la dependencia que ésta evolución tiene con los parámetros $\vec{\alpha}$ y con el punto inicial x_0 .

Un ejemplo de una ecuación diferencial discreta es la ecuación logística. Ésta modela la evolución en el número de individuos dentro de una población:

$$x_{n+1} = k x_n \left(1 - \frac{x_n}{M}\right)$$

Con k y M ambas positivas. Si consideramos x_n como la fracción de población máxima entonces ésta toma valores entre $0 \leq x_n \leq 1$ y la ecuación se convierte en:

$$x_{n+1} = \lambda x_n (1 - x_n) \quad (2.3)$$

Con $\lambda > 0$. De modo que depende únicamente de un parámetro λ y de la población inicial.

Análisis gráfico y cualitativo de sistemas

Para el análisis de la evolución de los sistemas, comúnmente se usa alguna de las herramientas: series de tiempo, diagrama fase o diagrama de bifurcación. Para los fines del presente informe basta con mencionar el diagrama fase aunque más detalle sobre la serie de tiempo y los diagramas de bifurcación puede encontrarse.

Diagramas fase

Ilustra los *saltos* que da el sistema entre estados consecutivos. Para esto se toma la sucesión de puntos $\mathcal{S} = \{x_0, x_1, \dots, x_n\}$ y se grafican las coordenadas $(x_0, x_1), (x_1, x_2), \dots, (x_{n-1}, x_n)$, es decir, en el eje de las x se grafica el punto del cual partió el sistema y en el eje de las y se grafica el punto al que *llegó* después de que se le aplicara la función f .

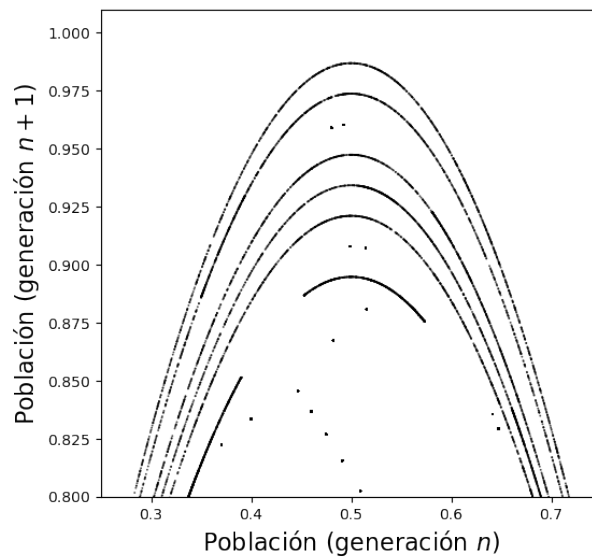


Figura 2.1: Diagrama fase que nos permite visualizar si el sistema se estabiliza. Básicamente nos dice de dónde proviene y a dónde llega el estado del sistema.

En la figura 2.1, cada arco representa la evolución de la ecuación logística 2.3 a un valor de λ . Es interesante notar que los arcos nunca se cruzan. Cada uno tiene una evolución diferente a las demás. Los pequeños arcos cerca del centro de la figura con sólo unos cuantos puntos (o incluso únicamente uno) representan los casos donde el sistema converge rápidamente a un punto o un ciclo límite.

Sistemas de Funciones iteradas y mapeo de Henón

Los sistemas de funciones iteradas son un tipo particular de ecuaciones diferenciales discretas en cuanto a que su estructura básica es la misma que en 2.2. Usualmente se les asocia con conjuntos autosemejantes aunque hay ciertas condiciones que deben cumplirse para que esto ocurra.

Unos de los sistemas de funciones iteradas más estudiados es el mapeo de Henón cuyas coordenadas a cada iteración se obtienen mediante las ecuaciones:

$$\begin{cases} x_{n+1} = 1 - a x_n^2 + y_n \\ y_{n+1} = b x_n \end{cases} \quad (2.4)$$

El atractor de este sistema para un millón de iteraciones se muestra en la figura 2.2.

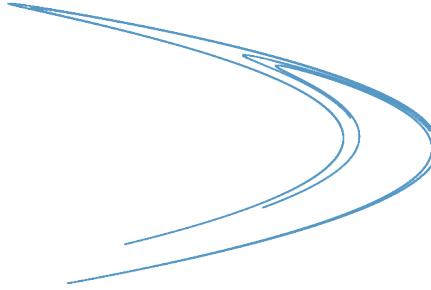


Figura 2.2: Atractor del conjunto de Henón cuyo sistema está representado por la ecuación 2.4

Juego del caos

El juego del caos, similar a los sistemas dinámicos discretos, consiste en un conjunto de condiciones iniciales y una función o acción a repetirse.

Uno de los más emblemáticos es el *Triángulo de Sierpinski* donde, sobre un triángulo equilátero con vértices $V = \{v_1 = (x_1, y_1), v_2 = (x_2, y_2), v_3 = (x_3, y_3)\}$ y etiquetados por $V_e = \{1, 2, 3\}$, se aplica el siguiente algoritmo:

1. Se elige aleatoriamente un número $\kappa \in V_e$ y se genera con ello el punto inicial ($p_0 = v_\kappa$) que esté entre los vértices de V .
2. Se elige aleatoriamente una nueva etiqueta $\kappa_1 \in V_e$ y se calcula el punto medio entre p_0 y v_{κ_1} . Este nuevo punto es p_1 , el correspondiente a la iteración 1. Más sucintamente, $p_1 = PM(p_0, v_{\kappa_1})$.
3. Se itera el paso anterior con una sutil pero importante diferencia: no se toma nuevamente p_0 para calcular el punto medio sino el punto generado durante la previa iteración. Es decir, $p_{n+1} = PM(p_n, v_{\kappa_{n+1}})$ para $n \geq 1$.

Se muestra en la figura 2.3a la gráfica de los puntos resultantes de cinco iteraciones; se puede inferir cuál fue la secuencia de etiquetas que los generó: $\{1, 3, 1, 2, 3, 2\}$. La gráfica en 2.3b se obtuvo con un millón de iteraciones y se puede decir que ningún punto cayó dentro del triángulo blanco invertido que está en el centro de la figura y que es semejante al original. Más aún, en 2.3c se cambiaron los colores de tres triángulos más pequeños para evidenciar que éstos, a su vez, tienen triángulos blancos invertidos todos ellos semejantes.

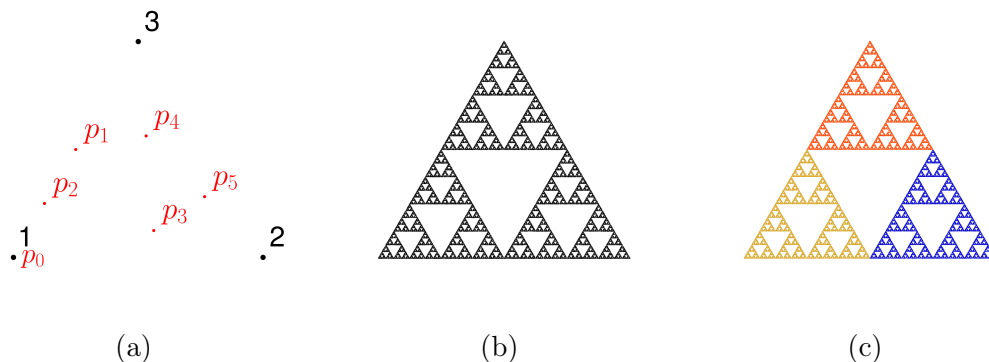


Figura 2.3: (a): Cinco puntos generados con el algoritmo del triángulo de Sierpinski. (b): Gráfica de los puntos después de un millón iteraciones. (c): Los tres triángulos pequeños reproducen la forma del original en (b), la gráfica es auto-semejante.

Juego del Caos en el DNA

El DNA yace dentro de cada célula de cada ser vivo. Tiene la información suficiente para que la célula cumpla con las todas funciones necesarias para sustentar la vida. Es una macromolécula que forma una espiral y en cuyo interior hay únicamente cuatro posibles moléculas, llamadas nucleótidos: Adenina, Guanina, Citosina, Timina. De ahora en adelante abreviaremos estos elementos con el diccionario: {Adenina: A, Guanina: G, Citosina: C, Timina: T}.

Para fines prácticos, consideraremos que el DNA es una larga cadena de caracteres donde a cada posición la ocupa alguno de los elementos del diccionario de arriba.

El juego del caos en el DNA consiste en:

- Dibujar un cuadrado donde a cada vértice se le etiqueta con uno de los elementos del diccionario A, T, C, G de modo que a cada vértice le corresponda únicamente un elemento del diccionario.
- Se identifica la letra que ocupa la primera posición dentro de la secuencia y se coloca un punto sobre el vértice que la representa en el cuadrado del paso anterior. A este punto inicial lo llamamos x_0 .
- Se identifica la letra que le sucede a la del punto anterior y se coloca un nuevo punto (x_1) en el punto medio entre x_0 y el vértice que representa dicha letra.

- Se itera la acción anterior. Se fija qué letra ocupa la posición n de la cadena y se traza el punto medio entre el punto x_{n-1} y el vértice que representa a la letra en la posición n .

Tomemos, a modo de ejemplo, la siguiente sub-cadena del DNA:

ATGCCGATA

En la figura 2.4 se ilustra cómo luce la gráfica del algoritmo para las primera 9 posiciones de la cadena.

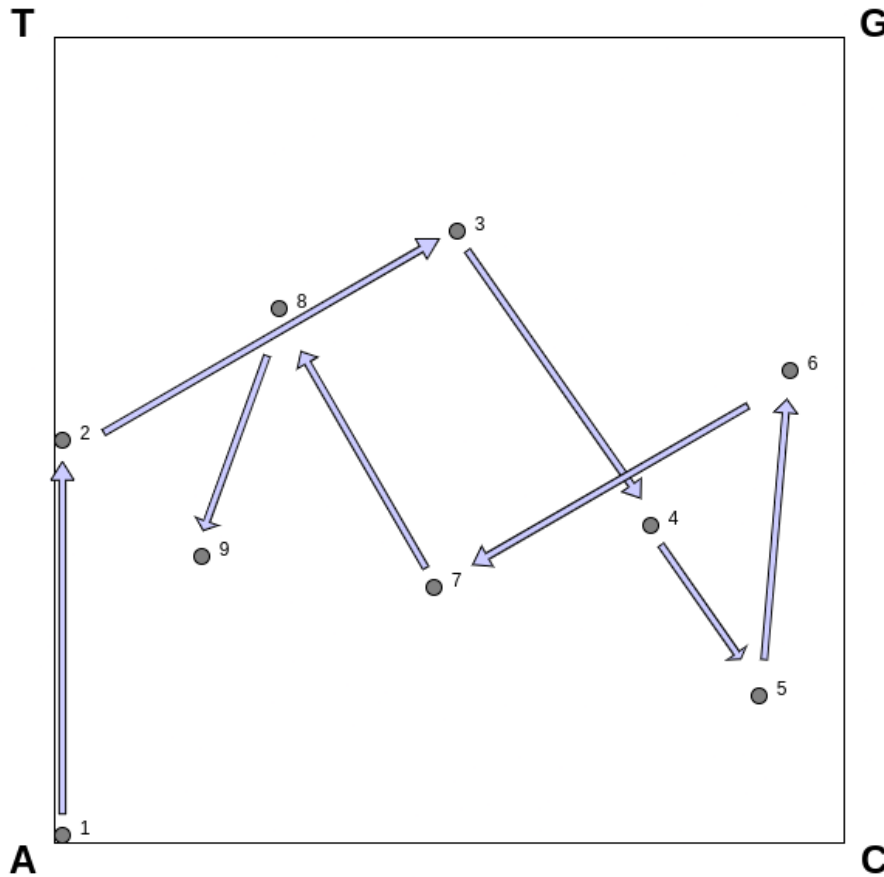


Figura 2.4: Algoritmo del juego del caos en secuencias de genomas. Se muestran los primeros nueve puntos del

Fractales

Los fractales son conjuntos cuya principal característica es que tienen una *dimensión fraccionaria*. Otra manera de decir esto requiere que pensemos en las dimensiones como el número de veces que un objeto se replica a sí mismo cuando se aumenta o se reduce uno de sus lados como una potencia de 2.

La dimensión se calcula con la expresión: $2^k = M$ donde k es la dimensión y M es el número de copias resultantes de duplicar el tamaño de un objeto. Acomodando un poco la expresión anterior llegamos a que $k = \ln(M)/\ln(2)$.

La figura 2.5 ilustra cómo esto se relaciona con la dimensión topológica. En 2.5a, duplicar el tamaño de un segmento de recta resulta en dos copias de la original, entonces $k = \ln(2)/\ln(2) = 1$. Cuando se lleva esta idea a un cuadrado como en 2.5b, aumentar el tamaño del cuadrado en un factor de 2 produce ahora cuatro copias de la figura inicial: $k = \ln(4)/\ln(2) = 2$. Finalmente en 2.5c se muestra únicamente la silueta de los tres triángulos de 2.3c. Duplicar las dimensiones del triángulo resulta en tres copias de éste, su dimensión es entonces $k = \ln(3)/\ln(2) \approx 1.584$.

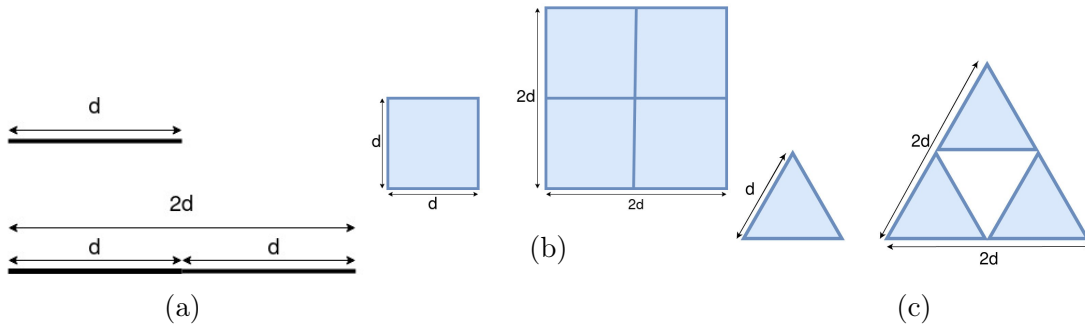


Figura 2.5: El cálculo de la dimensión fractal coincide con la topológica para una y dos dimensiones. El número de copias crece como potencia de 2 al duplicar la longitud de un objeto. Cuando el conjunto es auto-semejante y no llena el espacio; la dimensión fractal no es un número entero.

Una de las implicaciones que esto tiene es que un fractal no necesariamente *llena* el espacio donde se encuentra cuando se hacen copias de él, dejando espacios vacíos. La frontera del fractal suele describir vericuetos donde se puede verificar la autosemejanza, es decir, que el objeto se replica a sí mismo independientemente de la escala a la que se le observe.

Resulta sencillo calcular la dimensión fractal de conjuntos con evidente autosemejanza aunque esta cualidad no siempre aparezca durante el estudio de sistemas dinámicos. En la siguiente sección se introduce la dimensión de Minkowski-Bouligand, un método numérico para el cálculo de la dimensión que aproxima la dimensión de fractales.

Dimensión de Minkowski-Bouligand

Si bien la manera de calcular la dimensión fractal con la definición y el conteo del número de réplicas que se genera de la figura al duplicar el tamaño de uno de los lados; tiene limitaciones cuando la autosemejanza es no trivial o incluso inexistente.

Para calcular la dimensión fractal de conjuntos donde no sea posible contar el número de réplicas, se usa el algoritmo de *Minkowski-Bouligand* también conocido como el algoritmo de *conteo de cajas*. También se basa en la expresión $2^k = M$ y busca particionar el espacio métrico dentro del cual se encuentra el conjunto en segmentos cuyo tamaño decrezca como potencia de $\sim \frac{1}{2^n}$ y donde n es un número natural sobre el cual se itera.

A medida que se particione el espacio métrico, se cuenta el número de cajas dentro de las cuales haya al menos un elemento del conjunto. La dimensión se calcula con la expresión:

$$\dim_{box}(\mathcal{S}) = \lim_{\epsilon \rightarrow 0} \frac{\log(N(\epsilon))}{\log(1/\epsilon)} \quad (2.5)$$

Donde ϵ es la longitud de cada elemento de la partición y $N(\epsilon)$ es el número de cajas que contienen al menos un elemento del conjunto R .

En la figura 2.6 se muestran las primeras cuatro iteraciones del algoritmo del conteo de cajas aplicado sobre el triángulo de Sierpinski. Se aprecia cómo después de unas cuantas iteraciones, hay cuadrados dentro del triángulo invertido del centro donde no hay elementos del conjunto.

La manera de implementar la ecuación 2.5 para el cálculo de la dimensión consiste en iterar para diferentes valores escalas de ϵ y registrar $N(\epsilon)$. Luego se traza una recta con los valores de $\log(1/\epsilon)$ sobre el eje de las x y $\log(N(\epsilon))$ sobre el eje de las y .

Esto resulta en una recta cuya pendiente aproximará la dimensión fractal que deseamos calcular. De modo que se hace una regresión lineal con los puntos que resulten de la iteración para diferentes escalas de ϵ .

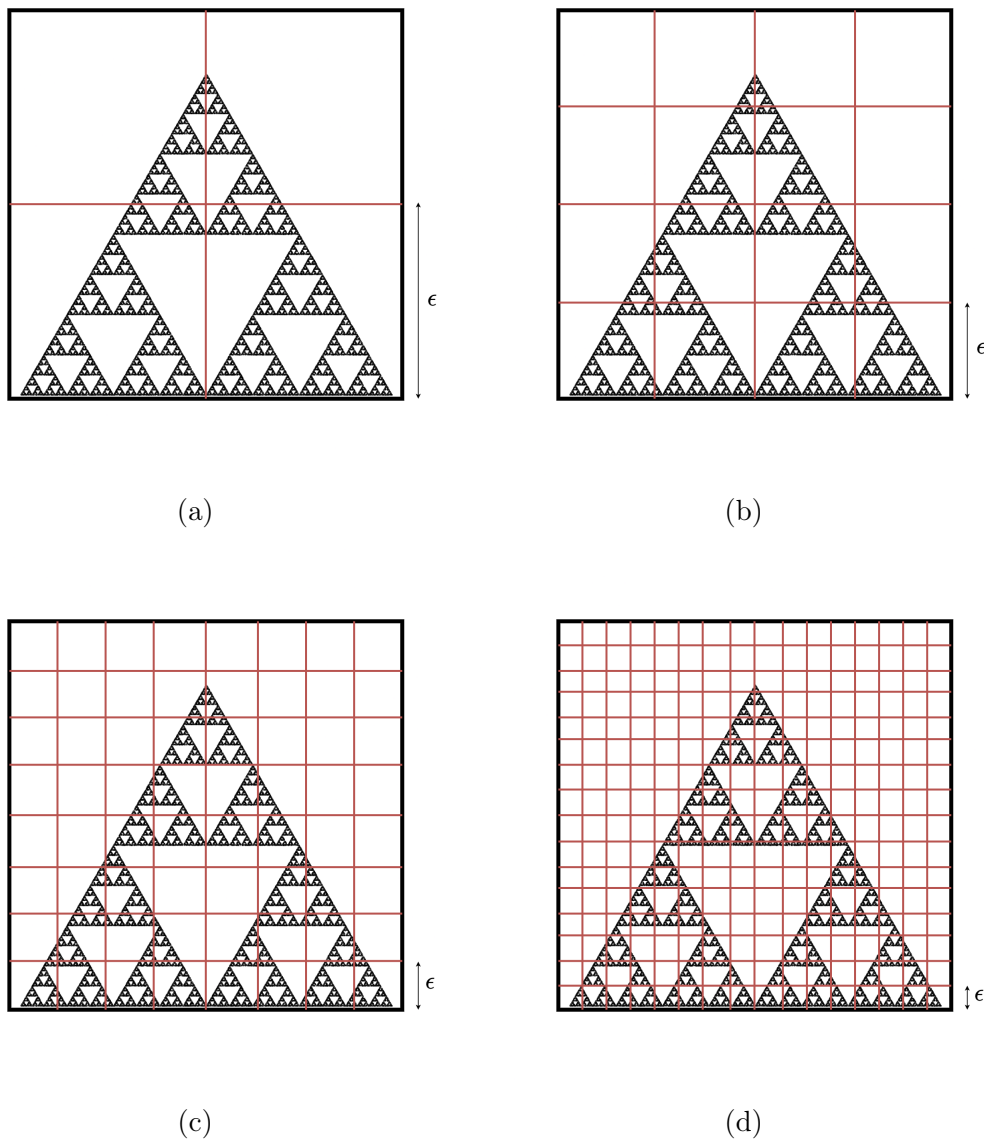


Figura 2.6: Particiones del espacio que contiene al conjunto. El número de particiones dentro de \mathbf{R}^2 incrementa como función de 2^n . A cada iteración se cuenta el número de *cuadrados* que tengan al menos un elemento del conjunto \mathcal{S} en su interior.

Capítulo 3

Implementación

Se desarrolló un proyecto en el lenguaje de programación python 3.x que calcula la dimensión de Minkowski-Bouligand de cualquier subconjunto de \mathbb{R}^2 dadas sus coordenadas. Se probó el algoritmo con los siguientes conjuntos:

- Una recta.
- Un cuadrado.
- Triángulo de Sierpinski.
- Conjunto de Henón
- Código de barras.
- Juego del caos en una subcadena del cromosoma 19.
- Juego del caos en una cadena aleatoria con el mismo diccionario del DNA.

Puede modificarse el número de elementos y extenderse para cualquier la dimensión de cualquier otro conjunto mediante la ejecución del método `Frac_dimension_computation()` incluido en la documentación del proyecto, mismo que puede encontrarse *dando click aquí*.

Se usó este desarrollo para reproducir el juego del caos en subcadenas del cromosoma 19 de longitud variable. También se generaron cadenas aleatorias a partir del diccionario de nucleótidos y se reprodujo el juego del caos en éstas también.

A cada uno de los conjuntos antes mencionados se le calculó la dimensión de Minkowski-Bouligand con el fin de indagar en la convergencia de la dimensión entre conjuntos aleatorios y conjuntos con cierto arreglo a partir del DNA real.

Las longitudes de la subcadena del cromosoma 19 que se usaron para el experimento fueron: $[10^4, 10^5, 10^6, 10^7, 10^8]$ y las longitudes de la cadena generada aleatoriamente fueron: $[10^4, 10^5, 10^6, 10^7]$. Esto se hizo así porque se observó que la convergencia en la dimensión para el caso de la cadena aleatoria siempre ocurría entre las potencias 10^5 y 10^6 . Además que el equipo de cómputo donde se desarrolló el proyecto tenía dificultades

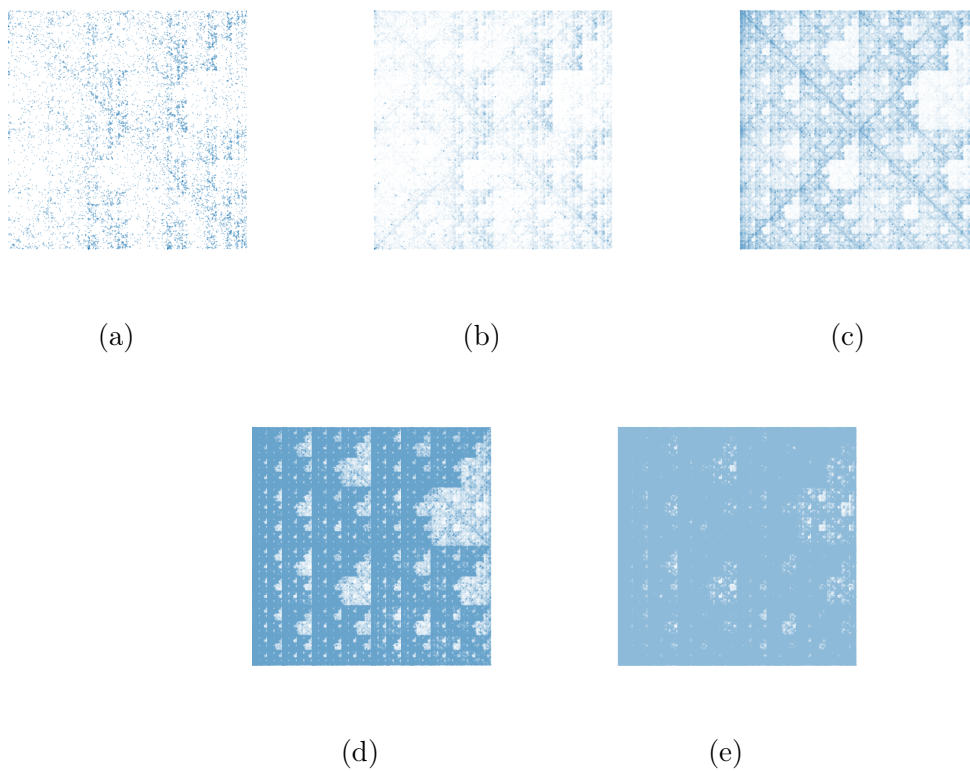


Figura 3.1: Gráficas del juego del caos aplicado a sub-cadenas del cromosoma 19.

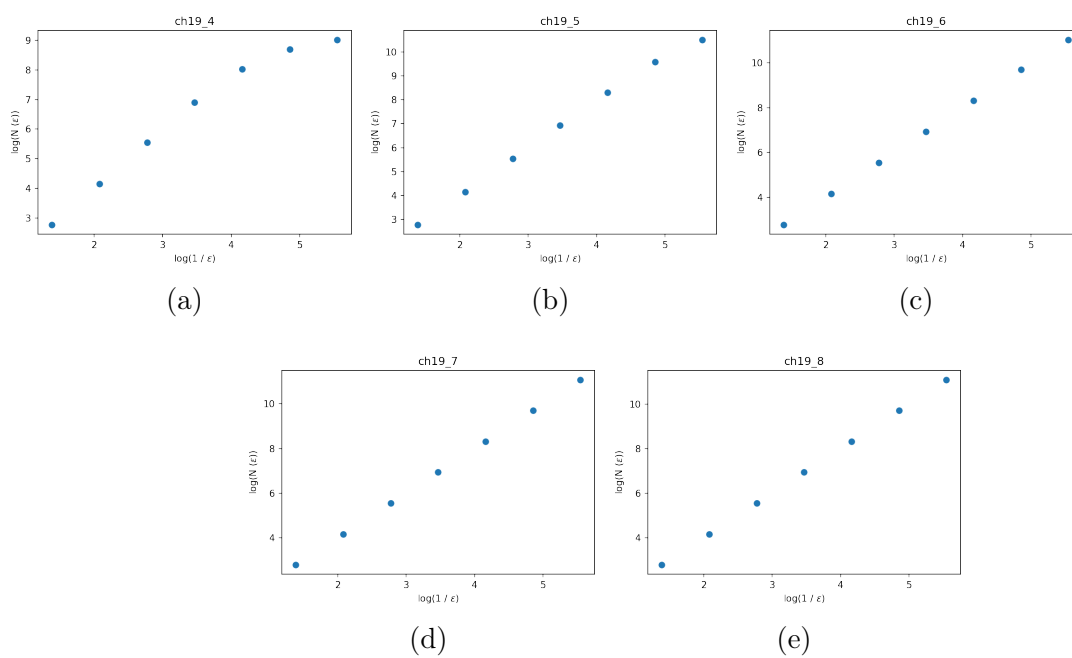


Figura 3.2: Gráficas del juego del caos aplicado a sub-cadenas del cromosoma 19.

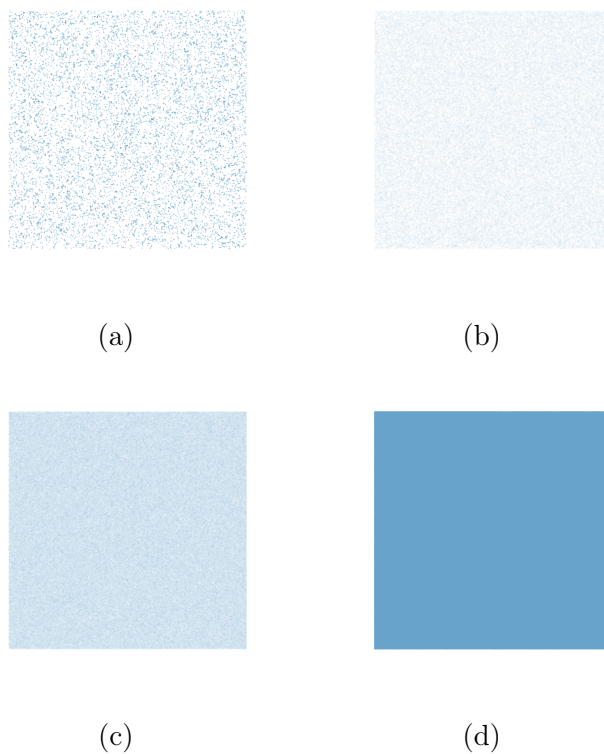


Figura 3.3: Gráficas del juego del caos en cadenas generadas aleatoriamente

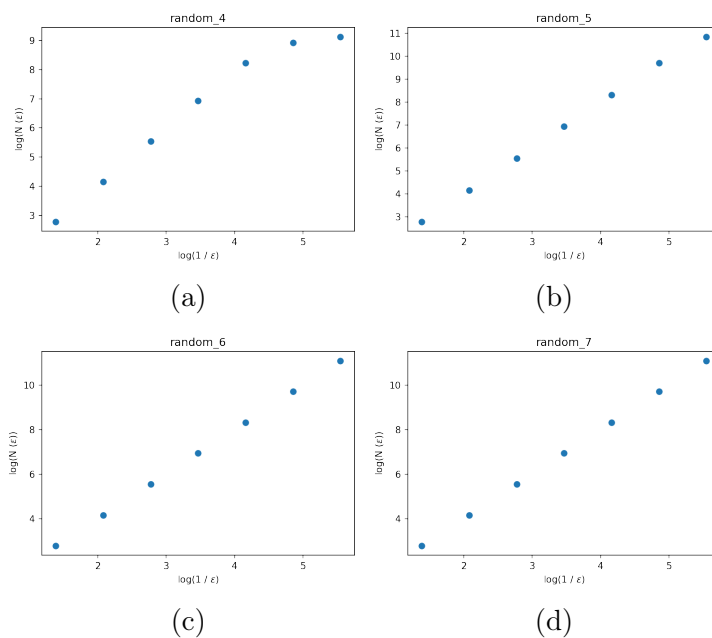


Figura 3.4: Gráficas de la recta que se forma por la relación de la ecuación 2.5 en el conjunto de cadenas de caracteres aleatorios.

Tabla 3.1: Dimensión de los conjuntos anteriormente descritos. *str* representa el origen de los datos y N se relaciona con la longitud de la cadena mediante la expresión: $len = 10^N$

N	str	chr_19	aleatoria
4		1.56	1.61
5		1.90	1.96
6		1.99	2.00
7		2.00	2.00
8		2.00	-

al generar la cadena de caracteres de longitud 10^8 .

En la tabla 3 se incluyen las dimensiones aproximadas de los conjuntos anteriormente descritos para diferentes longitudes. N se relaciona con la longitud de cada cadena según la ecuación: $len = 10^N$.

Se podría concluir que las dimensiones por conteo de cajas de ambos sistemas convergen a 2.0, el caso del juego del caos aplicado al DNA nunca llegó al valor de 2.0 sino que lo hizo después de que se redondeara a dos cifras después del punto decimal.

También es menester considerar que, a juzgar por las figuras de los atractores del sistema mostrados en la figura 3.1, es posible que para subcadenas del cromosoma de gran longitud el sistema se esté saturando y la dimensión no sea la correcta.

Para conocer acertadamente la dimensión de Minkowski-Bouligand, se calculó la dimensión para 500 subcadenas cuyas longitudes particionaran el intervalo $[10^5, 10^7]$ en 500 segmentos de igual tamaño. La gráfica de la longitud de la cadena vs la dimensión se muestra en la figura 3.5.

Notamos con esta figura que la, efectivamente, la dimensión se estabiliza rápidamente y se acerca asintóticamente a 2.0. Para determinar la dimensión real del sistema se propone:

- Reducir la muestra a únicamente 50 puntos a partir de 10^5 .
- Seleccionar únicamente los primeros 10 y los últimos 10 puntos de la gráfica.
- Hacer una regresión lineal para cada uno de los dos segmentos anteriormente mencionados.
- Resolver el sistema de ecuaciones con los valores de las pendientes e interceptos del paso anterior.

- Determinar el punto sobre la curva de dimensiones que más se aproxime a la intersección entre las rectas del paso anterior.

Los pasos que se enunciaron anteriormente se ilustran gráficamente en las figuras de 3.5. Lo que se busca con esto es aproximar el punto de inflexión de la curva de dimensiones. Consideraremos este valor como la mejor aproximación de la dimensión de Minkowski-Bouligand para el juego del caos aplicado al cromosoma 19.

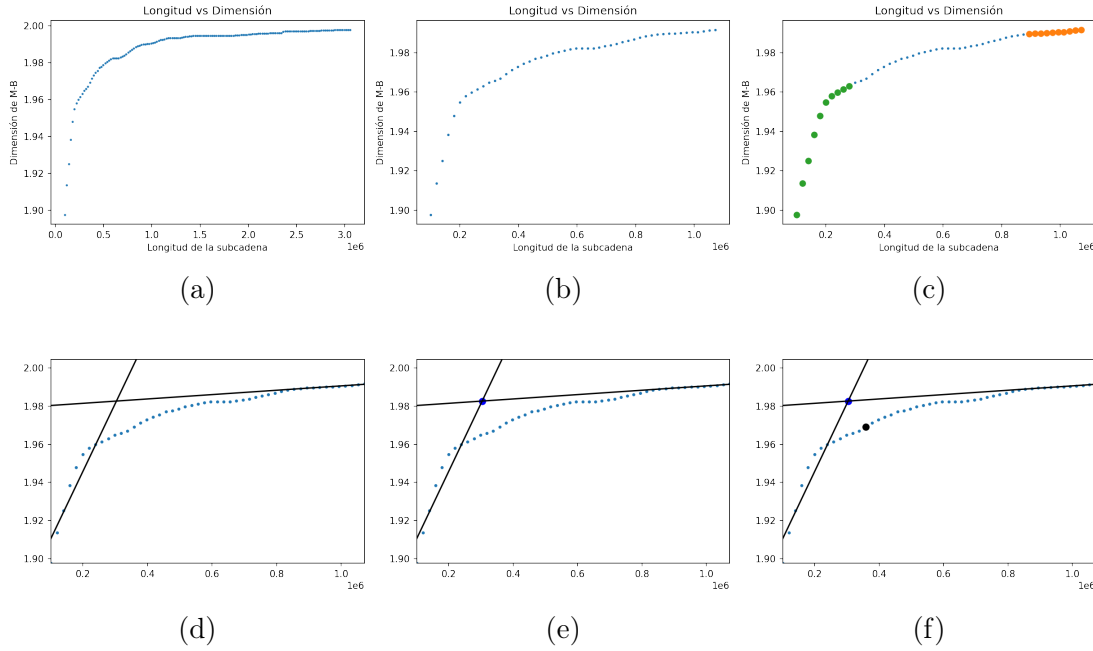


Figura 3.5: Ilustración de pasos para determinar la relación entre la longitud de la subcadena y la dimensión de Minkowski-Bouligand y así determinar el punto de inflexión que consideraremos la dimensión del atractor.

Esto condujo a identificar el punto indicado en la figura 3.5 y el valor aproximado de la dimensión de 1.97.

Discusión

De acuerdo con los resultados de la tabla 3, las dimensiones de Minkowski-Bouligand para juego del caos aplicado al cromosoma 19 y a cadenas aleatorias de caracteres parecieran converger ambos a 2.0. Aunque el análisis de la relación entre longitud de la subcadena y la dimensión condujo a un valor aproximado de la dimensión igual a 1.97.

El proyecto computacional completo con los métodos desarrollados para el cálculo de la dimensión así como varios ejemplos puede encontrarse *aquí*.

Bibliografía

- [1] Boeing, G. *Visual Analysis of Nonlinear Dynamical Systems: Chaos, Fractals, Self-Similarity and the Limits of Prediction*. Systems 2016, 4, 37.
- [2] M. Barnsley, A. Vince, *The chaos game on a general iterated function system*. *Ergodic Theory and Dynamical Systems*, 31(4), 1073-1079. doi:10.1017/S0143385710000428, (2011)
- [3] M. Barnsley, *Fractals Everywhere* 2nd ed. Academic Press, Atlanta, Georgia, (1993).
- [4] M. Schroeder *Fractals, chaos and power laws - Minutes from an Infinite Paradise* Freeman (1996).
- [5] M. Hirsch, S. Smale, R. Devaney *Differential Equations, Dynamical Systems and An Introduction to Chaos* 2nd ed. Elsevier Academic Press, 2004.
- [6] M. Newville, et al. *LMFIT: Non-Linear Least-Square Minimization and Curve-Fitting for Python*. Zenodo. 10.5281/zenodo.11813, 2014.