



UNIVERSIDAD AUTÓNOMA METROPOLITANA

DIVISIÓN DE CIENCIAS BÁSICAS E INGENIERÍA

DEPARTAMENTO DE FÍSICA

---

# Comportamiento Multifractal del Genoma Humano

---

*Tesis presentada por*

Yulián Andrés

ÁLVAREZ BALLESTEROS

*Director de tesis*

José Luis

DEL RÍO CORREA

*Subdirector de tesis*

Dr. Mario Alan

QUIROZ-JUAREZ

---

DOCTORADO EN CIENCIAS (FÍSICA)

*Para optar por el título de*

DOCTOR EN CIENCIAS (FÍSICA)

Ciudad de México, Noviembre de 2025

---



# Índice general

|                                                                           |            |
|---------------------------------------------------------------------------|------------|
| <b>Resumen</b>                                                            | <b>VII</b> |
| <b>Abstract</b>                                                           | <b>IX</b>  |
| <b>Introducción</b>                                                       | <b>XI</b>  |
| <br>                                                                      |            |
| <b>I Fundamentos Teóricos</b>                                             | <b>1</b>   |
| <br>                                                                      |            |
| <b>1. Bases teóricas para el análisis fractal y multifractal</b>          | <b>3</b>   |
| 1.1. Una aproximación topológica a los fractales . . . . .                | 4          |
| 1.1.1. Construyendo el espacio fractal . . . . .                          | 5          |
| 1.1.2. El papel de las transformaciones . . . . .                         | 6          |
| 1.2. Sistemas de Funciones Iteradas (IFS) . . . . .                       | 7          |
| 1.2.1. El triángulo de Sierpinski . . . . .                               | 7          |
| 1.3. Direcciones en un fractal . . . . .                                  | 9          |
| 1.4. Nociones de sistemas dinámicos . . . . .                             | 10         |
| 1.5. Dinámica sobre fractales . . . . .                                   | 10         |
| 1.6. La noción de dimensión fractal . . . . .                             | 10         |
| <br>                                                                      |            |
| <b>2. Fundamentos de las Medidas Invariantes en Estructuras Fractales</b> | <b>13</b>  |
| 2.1. Álgebras y $\sigma$ -Álgebras de Conjuntos . . . . .                 | 16         |
| 2.2. El Concepto de Medida . . . . .                                      | 18         |
| 2.3. Integrales y Métricas para Medidas . . . . .                         | 20         |
| 2.4. Densidades Invariantes en Sistemas Dinámicos . . . . .               | 22         |
| 2.4.1. Conjugación de Densidades . . . . .                                | 24         |
| 2.5. El Operador de Perron-Frobenius . . . . .                            | 25         |
| 2.5.1. Puntos Fijos del Operador . . . . .                                | 26         |
| 2.5.2. Espectro del Operador . . . . .                                    | 27         |
| 2.6. Operador de Koopman . . . . .                                        | 27         |

|           |                                                                                  |           |
|-----------|----------------------------------------------------------------------------------|-----------|
| 2.6.1.    | Propiedades del Operador de Koopman . . . . .                                    | 28        |
| 2.6.2.    | Espectro del Operador de Koopman . . . . .                                       | 29        |
| 2.6.3.    | Aplicación . . . . .                                                             | 29        |
| 2.6.4.    | Descomposición Modal Dinámica . . . . .                                          | 32        |
| 2.6.5.    | Relación con la Ergodicidad . . . . .                                            | 32        |
| 2.6.6.    | Ejemplo de Aplicación en Billares Clásicos . . . . .                             | 33        |
| 2.6.7.    | Comparación con Métodos Numéricos para Billares Cuánticos . . . . .              | 34        |
| 2.6.8.    | Ventajas Computacionales . . . . .                                               | 35        |
| 2.7.      | El Operador de Markov como Transformación Contractiva en el Espacio de Medidas . | 36        |
| 2.7.1.    | Interpretación Geométrica del Operador de Markov . . . . .                       | 37        |
| 2.7.2.    | Propiedades Básicas de la Medida Invariante . . . . .                            | 38        |
| 2.8.      | Teorema de Elton y sus Consecuencias . . . . .                                   | 39        |
| 2.9.      | Sistemas de Funciones Iteradas Recurrentes . . . . .                             | 40        |
| 2.9.1.    | Método del Juego del Caos para RIFS . . . . .                                    | 42        |
| 2.9.2.    | Método de la Máquina de Copiado . . . . .                                        | 44        |
| 2.9.3.    | RIFS y Secuencias de ADN . . . . .                                               | 47        |
| <b>3.</b> | <b>Multifractales</b>                                                            | <b>51</b> |
| 3.1.      | Medidas que Presentan Organización Multifractal . . . . .                        | 52        |
| 3.1.1.    | Medida Asociada al Conjunto de Cantor . . . . .                                  | 52        |
| 3.1.2.    | Fractales Autoafines de Soporte Continuo . . . . .                               | 55        |
| 3.1.3.    | La Dimensión de Rényi: Una Perspectiva desde la Teoría de la Información . .     | 59        |
| 3.1.4.    | Caracterización Multifractal mediante Dimensiones de Rényi . . . . .             | 61        |
| 3.1.5.    | Interrelación entre $D_q$ , $\tau(q)$ y la Autosimilitud . . . . .               | 63        |
| 3.2.      | El Espectro Multifractal $f(\alpha)$ y su Dualidad con $\tau(q)$ . . . . .       | 64        |
| 3.3.      | Turbulencia . . . . .                                                            | 65        |
| 3.4.      | Atractores Extraños en Sistemas Dinámicos Caóticos . . . . .                     | 68        |
| 3.5.      | La Medida Multinomial: Un Marco General . . . . .                                | 71        |
| 3.5.1.    | Teorema de Eggleston y Dimensiones Fractales . . . . .                           | 71        |
| 3.5.2.    | Visualización: Espectros Ópticos Multifractales . . . . .                        | 72        |
| 3.5.3.    | Comportamiento Local y Exponente de Hölder . . . . .                             | 72        |
| 3.5.4.    | Vínculos con la Mecánica Estadística de Boltzmann . . . . .                      | 74        |
| 3.5.5.    | El Conjunto de Condensación de una Medida . . . . .                              | 78        |
| 3.5.6.    | Reescalado con una Media Distorsionada: $P(q)$ . . . . .                         | 79        |
| 3.6.      | Caracterización mediante Momentos Globales y Dualidad de Legendre . . . . .      | 80        |
| 3.6.1.    | El Exponente de Correlación $\tau(q)$ como Momento Escalante . . . . .           | 82        |

|                                                                                                       |            |
|-------------------------------------------------------------------------------------------------------|------------|
| 3.6.2. La Transformada de Legendre como Puente entre lo Local y lo Global . . . . .                   | 83         |
| 3.7. Resumen y Conclusiones del Capítulo . . . . .                                                    | 86         |
| <b>II Aplicación al Genoma Humano</b>                                                                 | <b>87</b>  |
| <b>4. Representación Genómica Binaria</b>                                                             | <b>91</b>  |
| 4.1. Generación Fractal con el Juego del Caos . . . . .                                               | 91         |
| 4.1.1. Representación gráfica del ADN . . . . .                                                       | 92         |
| 4.2. Representación Genómica Binaria (BGR) . . . . .                                                  | 94         |
| 4.2.1. Ubicación de puntos en la órbita del $DW$ y sus puntos gráficos . . . . .                      | 96         |
| 4.3. Codificación de la información de la SG . . . . .                                                | 103        |
| 4.4. Datos y Metodología . . . . .                                                                    | 106        |
| 4.5. Resultados y Análisis . . . . .                                                                  | 107        |
| <b>5. Análisis Multifractal del Genoma Humano</b>                                                     | <b>117</b> |
| 5.0.1. Citogenética Humana . . . . .                                                                  | 117        |
| 5.0.2. Representaciones Gráficas de Secuencias Genómicas . . . . .                                    | 119        |
| 5.0.3. Formalismo Multifractal . . . . .                                                              | 126        |
| 5.0.4. Análisis Multifractal . . . . .                                                                | 128        |
| 5.1. Metodología . . . . .                                                                            | 133        |
| 5.2. Resultados y Análisis . . . . .                                                                  | 134        |
| 5.2.1. Citogenética del CGR . . . . .                                                                 | 141        |
| 5.2.2. Métodos de CM y BGR . . . . .                                                                  | 143        |
| 5.2.3. Ejemplo de Aplicación en los Cromosomas 9 e Y . . . . .                                        | 145        |
| <b>Conclusiones y perspectivas</b>                                                                    | <b>151</b> |
| <b>Bibliografía</b>                                                                                   | <b>154</b> |
| <b>Apéndice</b>                                                                                       | <b>159</b> |
| <b>A. Representación Genómica Binaria de los Cromosomas para Diferentes Valores de <math>M</math></b> | <b>161</b> |
| A.1. Comparación entre Cromosomas . . . . .                                                           | 161        |
| A.2. Representaciones Tridimensionales por Cromosoma . . . . .                                        | 174        |
| A.3. Espectros Multifractales . . . . .                                                               | 182        |
| <b>Glosario</b>                                                                                       | <b>183</b> |

---



# Resumen

La caracterización de la estructura del genoma humano representa un objetivo fundamental en genómica, con implicaciones para comprender los mecanismos de regulación génica, la organización nuclear y la relación entre la arquitectura del ADN y su función biológica. Los métodos tradicionales de análisis secuencial, aunque útiles, no capturan completamente la complejidad estructural y las propiedades multi-escala de la organización genómica. Esta limitación motivó el desarrollo de un marco integrador que combina representaciones fractales y análisis multifractal para caracterizar cuantitativamente la arquitectura genómica.

La investigación parte de la premisa de que el genoma humano presenta propiedades fractales que reflejan tanto restricciones evolutivas como requerimientos funcionales. En la primera parte de la tesis, se establecen los fundamentos matemáticos para el análisis fractal, incluyendo sistemas de funciones iteradas y formalismos de medida invariante. Se desarrolla además un marco multifractal con la definición de espectros multifractales adaptados específicamente para el análisis de secuencias biológicas a gran escala.

La necesidad de superar las limitaciones de precisión numérica y control de resolución en los métodos existentes llevó al desarrollo de la Representación Binaria Genómica como una adaptación paramétrica del Chaos Game Representation tradicional. Este enfoque implementa una codificación binaria que asigna cada nucleótido a un vector binario, eliminando los errores numéricos asociados a las operaciones en punto flotante. El algoritmo del Caminante Determinista Compactado genera trayectorias exactas en espacios discretos, donde un parámetro de resolución permite controlar explícitamente la escala espacial y realizar análisis a múltiples niveles. La reversibilidad del proceso garantiza que no se pierde información durante la transformación, mientras que los esquemas de codificación con vectores enteros reducen los requerimientos de almacenamiento y optimizan el procesamiento de genomas completos. Estas representaciones permiten caracterizar secuencias mediante análisis de frecuencias, mostrando conjuntos de secuencias cortas que comparten patrones de frecuencia. Estas distribuciones generan patrones fractales, particularmente a bajas frecuencias, que varían significativamente entre diferentes frecuencias y presentan localizaciones que alternan entre secciones repetitivas y no repetitivas de la secuencia.

La aplicación del formalismo multifractal al genoma humano completo permite caracterizar su

estructura heterogénea y compleja. El cálculo de espectros multifractales a nivel de genoma completo y por cromosoma individual proporciona un esquema que muestra de manera rápida, precisa y visual las diferencias entre estructuras. Los análisis revelan patrones distintivos que reflejan la organización estructural única de cada cromosoma, con variaciones significativas en los rangos de exponentes de Hölder y dimensiones fractales. El estudio detallado de los cromosomas 9 e Y muestra comportamientos multifractales atípicos, caracterizados por rangos excepcionalmente amplios de exponentes de Hölder que se correlacionan con su alto contenido en heterocromatina y regiones repetitivas. La detección de motivos repetitivos específicos en regiones heterocromáticas evidencia la relación entre arquitectura fractal y función genómica. La evaluación de la complementariedad entre aproximaciones geométrico-estructurales y estadísticas demuestra que, mientras los modelos de Markov reproducen estadísticas locales con bajo error promedio, el enfoque basado en representación binaria captura eficientemente la organización estructural global y los patrones en banda.

Los resultados proporcionan una caracterización de la organización estructural del genoma humano y establecen correlaciones entre parámetros fractales y elementos funcionales específicos, avanzando en la comprensión de los principios organizativos de la arquitectura genómica. La investigación concluye con contribuciones metodológicas en representación genómica, análisis multifractal y desarrollo de herramientas computacionales escalables, estableciendo bases para aplicaciones futuras en genómica médica, detección de variantes estructurales y estudios evolutivos comparativos.

# Abstract

The characterization of the human genome structure represents a fundamental objective in genomics, with implications for understanding gene regulatory mechanisms, nuclear organization, and the relationship between DNA architecture and its biological function. Traditional sequential analysis methods, although useful, do not fully capture the structural complexity and multi-scale properties of genomic organization. This limitation motivated the development of an integrative framework that combines fractal representations and multifractal analysis to quantitatively characterize genomic architecture.

The research is based on the premise that the human genome exhibits fractal properties that reflect both evolutionary constraints and functional requirements. In the first part of the thesis, the mathematical foundations for fractal analysis are established, including iterated function systems and invariant measure formalisms. Additionally, a multifractal framework is developed with the definition of multifractal spectra specifically adapted for the analysis of large-scale biological sequences.

The need to overcome limitations in numerical precision and resolution control in existing methods led to the development of Binary Genomic Representation as a parametric adaptation of traditional Chaos Game Representation. This approach implements binary encoding that assigns each nucleotide to a binary vector, eliminating numerical errors associated with floating-point operations. The Compact Deterministic Walker algorithm generates exact trajectories in discrete spaces, where a resolution parameter allows explicit control of spatial scale and enables analysis at multiple levels. Process reversibility guarantees no information loss during transformation, while integer vector encoding schemes reduce storage requirements and optimize processing of complete genomes. These representations enable sequence characterization through frequency analysis, showing sets of short sequences that share frequency patterns. These distributions generate fractal patterns, particularly at low frequencies, which vary significantly between different frequencies and present locations alternating between repetitive and non-repetitive sequence sections.

The application of multifractal formalism to the complete human genome enables characterization of its heterogeneous and complex structure. The calculation of multifractal spectra at both whole-genome and individual chromosome levels provides a scheme that quickly, accurately, and visually shows structural differences. Analyses reveal distinctive patterns reflecting the unique structural or-

ganization of each chromosome, with significant variations in Hölder exponent ranges and fractal dimensions. Detailed study of chromosomes 9 and Y shows atypical multifractal behaviors, characterized by exceptionally wide ranges of Hölder exponents that correlate with their high heterochromatin content and repetitive regions. The detection of specific repetitive motifs in heterochromatic regions provides evidence of the relationship between fractal architecture and genomic function. Evaluation of complementarity between geometric-structural and statistical approaches demonstrates that while Markov models reproduce local statistics with low average error, the binary representation-based approach efficiently captures global structural organization and banding patterns.

The results provide a characterization of the structural organization of the human genome and establish correlations between fractal parameters and specific functional elements, advancing the understanding of organizational principles governing genomic architecture. The research concludes with methodological contributions in genomic representation, multifractal analysis, and development of scalable computational tools, establishing foundations for future applications in medical genomics, structural variant detection, and comparative evolutionary studies.



# Introducción

La geometría fractal [1–7] ha mostrado ser una herramienta poderosa para describir estructuras complejas y autosimilares presentes en sistemas naturales. La representación del juego del caos (CGR), propuesta por Jeffrey [6] para el análisis de secuencias genómicas, permite mapear secuencias de nucleótidos a fractales en el espacio bidimensional. Sin embargo, este enfoque presenta limitaciones significativas en precisión numérica que impiden la reconstrucción exacta de secuencias y no permite un control explícito sobre la resolución de análisis.

El análisis computacional de secuencias genómicas enfrenta actualmente desafíos fundamentales derivados de la escala masiva de datos genómicos modernos. Con el genoma humano completo T2T-CHM13v2.0 [8] ocupando aproximadamente 3.1Gb de información, los métodos tradicionales de representación y análisis muestran limitaciones significativas en precisión numérica, escalabilidad y capacidad de caracterización estructural.

De manera paralela, la teoría multifractal [9–11] proporciona un marco sofisticado para caracterizar distribuciones heterogéneas complejas mediante espectros de dimensiones fractales generalizadas en lugar de una única dimensión. Como aplicación, esta teoría comienza a gestarse en el estudio de la distribución de energía en flujos turbulentos, tratados por Andrey Kolmogorov [12], a lo cual Benoit Mandelbrot [13] añade nuevas consideraciones, en donde de manera conjunta la escala (tamaño de los remolinos) y la energía (disipación), permitiendo que esta última se distribuya de forma distinta incluso dentro de una misma escala. Esta perspectiva se aproxima a la teoría multifractal, formalizada posteriormente por Uriel Frisch y Giorgio Parisi [14], quienes introducen el concepto de modelo multifractal. Desde entonces, este enfoque ha sido ampliamente utilizado para describir la distribución de energía en turbulencia [15], encontrando aplicaciones en diversas disciplinas, desde la meteorología hasta la economía. No obstante, su aplicación sistemática al análisis integral del genoma humano aún no ha sido explorada de manera exhaustiva.

La combinación de estas dos herramientas matemáticas fundamentales motiva este trabajo: desarrollar una reformulación del CGR que supere sus limitaciones numéricas mediante la Representación Binaria Genómica (BGR) [16–18], e integrar esta reformulación con análisis multifractal riguroso para proporcionar una caracterización cuantitativa más detallada de la organización estructural del genoma humano.

## Contexto científico

Los avances en genómica computacional han transitado desde métodos secuenciales básicos hacia enfoques integradores que combinan múltiples perspectivas analíticas. La reciente culminación del secuenciamiento del genoma humano T2T-CHM13v2.0 [8], que ha resuelto regiones previamente incompletas como centrómeros y telómeros [19, 20], proporciona la posibilidad de realizar análisis exhaustivos de la organización genómica desde una perspectiva estructural y fractal.

Las representaciones gráficas de secuencias genómicas, particularmente la Representación CGR [6], han demostrado capacidad para revelar patrones estructurales fundamentales que actúan como firmas características de las secuencias [21, 22]. Estas representaciones exponen una organización fractal y autosimilar intrínseca al material genético, donde la recurrencia de patrones a diferentes escalas constituye una forma de caracterización estructural única para cada región genómica. Específicamente, el análisis de estas representaciones permite identificar:

- Firmas fractales distintivas que diferencian regiones codificantes de no codificantes [21]
- Patrones autosimilares característicos en islas CpG/GpC asociadas con regulación epigenética [19]
- Estructuras repetitivas con propiedades de escalamiento específicas en centrómeros, telómeros y regiones heterocromáticas [3, 19, 20]
- Huellas fractales particulares en polimorfismos y variaciones estructurales [23]

Estas características fractales no solo reflejan la presencia diferencial de subsecuencias específicas en distintas regiones del genoma, sino que también constituyen firmas estructurales que hasta ahora no habían sido sistemáticamente asociadas con localizaciones genómicas específicas. La distribución heterogénea de motivos secuenciales, evidenciada por la predominancia estadística de ciertas subsecuencias en regiones particulares, encuentra en el análisis fractal un marco unificador que permite correlacionar estas propiedades estadísticas con características estructurales medibles.

Sin embargo, las limitaciones inherentes al CGR en cuanto a precisión en cálculos iterativos e imposibilidad de reconstrucción exacta han obstaculizado la explotación completa de estas firmas fractales para el análisis de genomas complejos. Similarmente, aunque la teoría multifractal [24, 25] proporciona herramientas robustas para caracterizar heterogeneidades estructurales, su aplicación sistemática requeriría desarrollos metodológicos específicos que permitieran asociar formalmente los espectros multifractales con regiones genómicas concretas y sus propiedades estadísticas subyacentes.

# Objetivos

## Objetivo general

Desarrollar y validar una metodología integral basada en la Representación Binaria Genómica y el análisis multifractal para la caracterización estructural del genoma humano, superando limitaciones de enfoques previos mediante adaptaciones rigurosas, paramétricas y aplicaciones computacionales eficientes.

## Objetivos específicos

- Establecer los fundamentos matemáticos rigurosos de espacios métricos, IFS, medidas invariantes y teoría multifractal aplicables a secuencias genómicas [2, 25, 26].
- Desarrollar el método de Representación Binaria Genómica como una adaptación del CGR que elimine limitaciones numéricas y permita reconstrucción exacta de secuencias [16, 17].
- Introducir el parámetro  $M$  en el Caminante Determinista Compactado para proporcionar control explícito sobre resolución y escala [18].
- Establecer equivalencia formal entre esquemas previos mediante BGR y desarrollar esquemas de codificación eficientes [16].
- Aplicar análisis multifractal sistemático al genoma humano completo mediante técnicas de box-counting, calculando espectros completos de singularidad y multifractal [11, 24].
- Caracterizar patrones multifractales distintivos en cromosomas individuales e identificar comportamientos relevantes en cromosomas específicos.
- Comparar complementariedades analíticas entre representaciones BGR y Cadenas de Markov [4] para diferentes tipos de estudios genómicos.
- Identificar y caracterizar secuencias repetitivas específicas, motivos complejos y variaciones estructurales en regiones como heterocromatina, centrómeros y telómeros, mediante análisis de frecuencias y secuenciación.
- Demostrar aplicabilidad de métodos desarrollados a genomas de diferentes longitudes y complejidades.
- Establecer bases para futuras investigaciones en análisis comparativo entre especies y citogenética.

Esta tesis está estructurada en dos partes principales, cada una con objetivos y metodologías diferenciadas:

## Parte I: Fundamentos Teóricos

Esta primera parte establece el marco matemático necesario para desarrollar los métodos presentados en la segunda parte. A pesar de que este material se encuentra mayormente desarrollado en la tesis de maestría *Sistemas de funciones iteradas, fractales y multifractales* [7], la inclusión de estos fundamentos busca proporcionar al lector las referencias matemáticas requeridas para comprender los desarrollos posteriores, sin necesidad de consultar fuentes externas. Además, permite contextualizar el enfoque adoptado en esta investigación.

**Capítulo 1 - Fundamentos Matemáticos:** Desarrolla los principios topológicos básicos requeridos para definir los sistemas de funciones iteradas simples y el operador de Hutchinson [2, 26]. El capítulo establece las bases conceptuales para entender la estructura geométrica de los fractales y su generación mediante transformaciones contractivas.

**Capítulo 2 - Introducción a la Medida Invariante sobre Fractales:** Define los sistemas de funciones iteradas con probabilidades y el algoritmo de iteración aleatoria [2]. Introduce el concepto de medida invariante como distribución de probabilidad sobre atractores fractales [25]. Desarrolla los elementos de teoría de medida necesarios para el análisis de estructuras fractales [27].

**Capítulo 3 - Multifractales:** Desarrolla la teoría multifractal partiendo de la medida de Cantor, introduciendo los conceptos de exponente de singularidad, espectros multifractales y dimensión prefractal [9–11]. Presenta la función de Cantor y demuestra la estructura multifractal de conjuntos autosimilares [3].

Para lectores con experiencia en teoría multifractal y sistemas de funciones iteradas, es posible proceder directamente a la Parte II, donde se desarrollan las aplicaciones al genoma humano. Se sugiere sin embargo una revisión de esta primera parte para conocer el enfoque específico adoptado en esta investigación.

## Parte II: Aplicación al Genoma Humano

La segunda parte aplica los fundamentos teóricos presentados anteriormente al análisis del genoma humano, desarrollando metodologías específicas y presentando resultados detallados.

**Capítulo 4 - Representación Genómica Binaria:** Presenta la representación genómica binaria (BGR) como base para las representaciones gráficas de secuencias. Examina el CGR tradicional [6] y sus limitaciones en precisión numérica, particularmente la acumulación de errores en operaciones de punto flotante y la imposibilidad de reconstrucción exacta de secuencias. Desarrolla la BGR [16, 17], que asocia cada nucleótido con vectores binarios específicos:  $A=(0,0)$ ,  $C=(0,1)$ ,  $G=(1,1)$ ,  $T=(1,0)$ .

Esta adaptación permite interpretar el CGR como la trayectoria de un caminante determinista en lugar de un proceso estocástico. Introduce el Caminante Determinista Compactado con parámetro  $M$  [18], que controla la resolución espacial en potencias de dos ( $2^M \times 2^M$  puntos). Establece la equivalencia formal entre los esquemas anteriores mediante la BGR y desarrolla métodos de codificación basados en vectores enteros que reducen los requerimientos de almacenamiento. Implementa algoritmos de codificación y decodificación que garantizan la reversibilidad completa de las representaciones. Genera los histogramas de frecuencias de subsecuencias para el genoma humano completo T2T-CHM13v2.0 [8], mostrando distribuciones fractales que varían significativamente entre diferentes frecuencias y localizaciones en la secuencia.

**Capítulo 5 - Análisis Multifractal del Genoma Humano:** Describe el contexto citogenético humano, incluyendo la estructura cromosómica, bandas citogenéticas y regiones funcionales como centrómeros, telómeros, heterocromatina e islas CpG. Aplica técnicas de box-counting [2] al genoma humano completo T2T-CHM13v2.0 [8], utilizando diferentes tamaños de caja para caracterizar la estructura a múltiples escalas. Calcula espectros multifractales completos ( $f(\alpha)$ ,  $\tau(q)$ ,  $D_q$ ) para el genoma completo y para cada cromosoma individual [11, 24]. Identifica patrones característicos y comportamientos particulares en cromosomas específicos, con especial atención a los cromosomas 9 e Y que muestran rangos más amplios de exponentes de Hölder. Compara cuantitativamente las representaciones BGR con modelos de Cadenas de Markov [4], mostrando que mientras los modelos de Markov reproducen estadísticas locales con error promedio del 2 % para cadenas de 12 nucleótidos, BGR captura mejor la organización estructural global. Identifica secuencias repetitivas específicas en regiones heterocromáticas, incluyendo motivos como AATGGAATGGA y TTCCATTCCATT, así como pentámeros ATTC-C/GGAAT asociados a proteínas de heterocromatina. Localiza polimorfismos y variaciones estructurales mediante análisis geométrico de los mapas BGR, estableciendo correlaciones entre parámetros fractales y características biológicas específicas.

El desarrollo de la Representación Binaria Genómica (BGR) y su combinación con análisis multifractal constituye un enfoque metodológico para la caracterización estructural del genoma humano. Esta propuesta aborda limitaciones de métodos anteriores mediante una reformulación matemática que posibilita la reconstrucción de secuencias y permite ajustar la resolución analítica mediante el parámetro  $M$ , alternativo al uso de ventanas deslizantes.

El estudio establece la equivalencia entre métodos anteriores mediante BGR, con el desarrollo de esquemas de codificación asociados. La aplicación de teoría multifractal al genoma humano identificó patrones estructurales a nivel cromosómico, con variaciones en cromosomas específicos que se relacionan con regiones como la heterocromatina. El análisis comparativo con modelos de Cadenas de Markov indicó que los enfoques geométricos y estadísticos proporcionan perspectivas complementarias para el estudio genómico.

Entre los aspectos desarrollados se incluye la adaptación a representaciones binarias, el control

de resolución mediante el parámetro  $M$ , y la identificación de elementos genómicos mediante análisis geométrico. La metodología resultante integra representaciones secuenciales con caracterización cuantitativa multiescala.

Las posibles aplicaciones de este trabajo comprenden áreas como la genómica computacional, donde podría contribuir al análisis de genomas de tamaño considerable; la genómica médica, mediante la identificación de variantes estructurales; y la biología evolutiva, a través de estudios comparativos de arquitecturas genómicas. Asimismo, podría servir como base para el desarrollo de herramientas bioinformáticas y para investigaciones sobre organización de información genética.

Este enfoque integrado ofrece perspectivas para investigaciones futuras en genómica comparada, caracterización de variantes estructurales y estudios de evolución genómica, dentro del campo más amplio del análisis cuantitativo de organización genómica.

# Parte I

## Fundamentos Teóricos





# Capítulo 1

## Bases teóricas para el análisis fractal y multifractal

Este capítulo introductorio pretende establecer los principios matemáticos elementales empleados en el marco teórico fractal y multifractal, temas que ya fueron abordados en la investigación de maestría [7] previa a este doctorado. Este enfoque de continuidad responde al lineamiento institucional de la Universidad Autónoma Metropolitana, particularmente de su departamento de física, que concibe el trabajo doctoral como una extensión natural del de maestría. Partiendo de esta premisa, se presenta una revisión de los conceptos fundamentales del análisis fractal y multifractal, apoyándose en el desarrollo teórico previamente discutido. Cabe aclarar que este cuerpo teórico y sus aplicaciones fueron contruidos por múltiples investigadores en décadas recientes. El propósito de este texto, al igual que el del trabajo antecedente, no es realizar una crónica histórica exhaustiva del campo, aunque en varios puntos se reconoce la autoría de contribuciones relevantes. El objetivo central, en cambio, es suministrar una base conceptual sólida para la comprensión de las nociones e instrumentos matemáticos clave que sustentan el análisis fractal y multifractal, orientados particularmente hacia su uso en el estudio del genoma humano.

Aquí se exponen las ideas elementales del análisis fractal, sentando los cimientos matemáticos de esta teoría. Se inicia con el concepto de espacios métricos para cuantificar distancias, crucial para caracterizar geometrías fractales. Se desarrollan las nociones de completitud y convergencia, indispensables para entender la estructura de los fractales. Se presenta el espacio  $\mathcal{H}(\mathbf{X})$  de fractales equipado con la métrica de Hausdorff, que permite analizar conjuntos complejos como entidades únicas. Se examinan las transformaciones contractivas y el teorema del punto fijo de Banach, instrumentos centrales para generar fractales mediante sistemas de funciones iteradas (IFS). Finalmente, se establecen las bases para comprender la dimensión fractal, una propiedad distintiva de estos objetos frente a las figuras geométricas tradicionales. Estos fundamentos proporcionan el lenguaje matemático necesario para adentrarse en el mundo fractal.

## 1.1. Una aproximación topológica a los fractales

Un enfoque inicial para concebir un fractal es visualizarlo como un subconjunto de un espacio matemático. Un espacio  $\mathbf{X}$  es simplemente un conjunto, y a sus elementos los llamamos *puntos*. Este espacio constituye el “escenario” donde existen los fractales. La estructura del espacio determina las propiedades que deben cumplir sus puntos. Por ejemplo, el conjunto de los números reales  $\mathbf{X} = \mathbb{R}$  es un espacio donde cada punto  $x \in \mathbf{X}$  representa un número real o una ubicación sobre una línea recta. De manera similar, el plano euclidiano  $\mathbf{X} = \mathbb{R}^2$  representa pares ordenados  $x = (x_1, x_2)$ . Posteriormente veremos cómo ciertas representaciones genómicas corresponden a subconjuntos de espacios planos. Los objetos de interés —números, imágenes, transformaciones, etc.—, llamados aquí puntos, deben pertenecer a un conjunto donde sea posible definir una noción de distancia entre cualquier par. Además, este conjunto debe poseer una propiedad de “saturación” que garantice, bajo ciertas condiciones, que las sucesiones tengan límites dentro del propio conjunto.

En el espacio  $\mathbf{X} = \mathbb{R}$  pueden definirse diversas métricas. Algunos ejemplos de distancia entre puntos  $x, y \in \mathbf{X}$  son:  $d(x, y) = |x - y|$  (euclidiana),  $d(x, y) = 2 \cdot |x - y|$ , o  $d(x, y) = |x^3 - y^3|$ , cuya elección depende del problema concreto. En el plano  $\mathbf{X} = \mathbb{R}^2$ , para puntos  $P = (x, y)$  y  $Q = (u, v)$ , también existen métricas comunes:  $d_2(P, Q) = \sqrt{(x - u)^2 + (y - v)^2}$  (euclidiana),  $d_\infty(P, Q) = \max\{|x - u|, |y - v|\}$  (del máximo),  $d_1(P, Q) = |x - u| + |y - v|$  (de Manhattan o taxista), cada una útil en contextos específicos. Dos métricas  $d_1$  y  $d_2$  sobre un mismo espacio  $\mathbf{X}$  se consideran *equivalentes* si existen constantes positivas  $0 < c_1 < c_2 < \infty$  cumpliendo

$$c_1 d_1(x, y) \leq d_2(x, y) \leq c_2 d_1(x, y) \quad \text{para todo } (x, y) \in \mathbf{X} \times \mathbf{X}.$$

La intuición es que dos métricas equivalentes proporcionan la misma percepción de proximidad, como si el espacio sufriera una deformación controlada donde las relaciones de distancia se preservan esencialmente. Dos espacios métricos  $(\mathbf{X}_1, d_1)$  y  $(\mathbf{X}_2, d_2)$  son *métricamente equivalentes* si existe una biyección  $h : \mathbf{X}_1 \rightarrow \mathbf{X}_2$  tal que la métrica inducida  $\tilde{d}_1(x, y) = d_2(h(x), h(y))$  sea equivalente a  $d_1$ . Esto implica que los espacios se relacionan mediante una transformación que no permite compresiones/expansiones ilimitadas, ni superposiciones o desgarres.

Una función  $f : \mathbf{X}_1 \rightarrow \mathbf{X}_2$  entre espacios métricos es *continua* si, para cualquier  $\epsilon > 0$  y cada  $x \in \mathbf{X}_1$ , podemos hallar  $\delta > 0$  tal que

$$d_1(x, y) < \delta \quad \text{implica} \quad d_2(f(x), f(y)) < \epsilon.$$

Si  $f$  es biyectiva y su inversa también es continua, entonces es un *homeomorfismo* y decimos que  $\mathbf{X}_1$  y  $\mathbf{X}_2$  son *homeomorfos*. La equivalencia métrica es una condición más fuerte que el homeomorfismo; esta última sólo preserva propiedades topológicas. Dos métricas sobre un mismo  $\mathbf{X}$  son topológicamente equivalentes si la función identidad es un homeomorfismo. Se puede probar que la equivalencia métrica implica la equivalencia topológica.

Tras definir espacios métricos, es necesario abordar límites de sucesiones y otras nociones para describir conjuntos. Una sucesión  $\{x_n\}$  en  $(\mathbf{X}, d)$  es una *sucesión de Cauchy* si, para cada  $\epsilon > 0$ , existe  $N$  tal que  $d(x_n, x_m) < \epsilon$  para todo  $n, m > N$ . Aunque sus términos se aproximan entre sí, esto no garantiza que converjan a un punto del espacio. Decimos que  $\{x_n\}$  *converge* a  $x \in \mathbf{X}$  si, dado  $\epsilon > 0$ , existe  $N$  con  $d(x_n, x) < \epsilon$  para  $n > N$ . Escribimos  $x = \lim_{n \rightarrow \infty} x_n$ . La bola cerrada  $\mathbf{B}(x, \epsilon) = \{y : d(x, y) \leq \epsilon\}$  contiene casi todos los términos de la sucesión convergente. Toda sucesión convergente es de Cauchy.

Un espacio métrico  $(\mathbf{X}, d)$  es *completo* si cada sucesión de Cauchy en  $\mathbf{X}$  converge a un punto de  $\mathbf{X}$ . El espacio  $(\mathbb{R}, d_{\text{Euclidiana}})$  es completo.

Para  $S \subset \mathbf{X}$ , un punto  $x \in \mathbf{X}$  es *punto límite* de  $S$  si existe una sucesión en  $S \setminus \{x\}$  que tiende a  $x$ . La *clausura*  $\bar{S}$  es  $S$  unida a sus puntos límite.  $S$  es *cerrado* si  $S = \bar{S}$ .  $S$  es *perfecto* si coincide con su conjunto de puntos límite. Si  $(S, d)$  es completo, entonces es cerrado. Recíprocamente, si  $\mathbf{X}$  es completo y  $S$  es cerrado, entonces  $(S, d)$  es completo.

$S$  es *compacto* si toda sucesión infinita en  $S$  posee una subsucesión convergente en  $S$ .  $S$  es *acotado* si está contenido en alguna bola.  $S$  es *totalmente acotado* si, para cada  $\epsilon > 0$ , puede cubrirse con un número finito de bolas de radio  $\epsilon$ . En un espacio completo,  $S$  es compacto si y solo si es cerrado y totalmente acotado. En  $\mathbb{R}^2$ , un conjunto es compacto precisamente cuando es cerrado y acotado.  $S$  es *abierto* si para cada  $x \in S$  existe  $\epsilon > 0$  con  $B(x, \epsilon) \subset S$ . El propio espacio  $\mathbf{X}$  es abierto y cerrado.

Todo subconjunto acotado de  $(\mathbb{R}^2, d_{\text{Euclidiana}})$  cumple la propiedad de Bolzano-Weierstrass, por lo que sus subconjuntos cerrados y acotados son compactos. Un punto  $x$  está en la *frontera*  $\partial S$  si toda bola centrada en  $x$  intersecta tanto a  $S$  como a su complemento.

$(\mathbf{X}, d)$  es *conexo* si sus únicos subconjuntos simultáneamente abiertos y cerrados son  $\mathbf{X}$  y  $\emptyset$ . Un subconjunto es conexo si lo es como subespacio. Es *completamente desconexo* si sus únicos subconjuntos conexos no vacíos son los unitarios.  $S$  es *conexo por caminos* si para cualquier par de puntos existe una función continua (camino) dentro de  $S$  que los une.

### 1.1.1. Construyendo el espacio fractal

Dado un espacio métrico completo  $(\mathbf{X}, d)$ , denotamos por  $\mathcal{H}(\mathbf{X})$  la colección de todos los subconjuntos compactos no vacíos de  $\mathbf{X}$ . Un elemento de  $\mathcal{H}(\mathbf{X})$  puede consistir en uno o varios de tales conjuntos. Para  $\mathbf{X} = \mathbb{R}^2$ , un punto en  $\mathcal{H}(\mathbb{R}^2)$  puede interpretarse como una imagen. Para  $x \in \mathbf{X}$  y  $B \in \mathcal{H}(\mathbf{X})$ , definimos la *distancia de un punto a un conjunto* como  $d(x, B) = \min\{d(x, y) : y \in B\}$ .

Con la métrica euclidiana, los puntos a distancia fija de un punto dado forman una circunferencia. Si  $B = \{x\}$  es un único punto, las curvas de nivel de  $d(\cdot, B)$  son circunferencias.

Para  $A, B \in \mathcal{H}(\mathbf{X})$ , definimos  $D(A, B) = \max\{d(x, B) : x \in A\}$ , la *distancia de A a B*. En general  $D(A, B) \neq D(B, A)$ , por lo que  $D$  no es métrica. La *distancia de Hausdorff* es  $h(A, B) =$

$\max\{D(A, B), D(B, A)\}$ . Esta sí define una métrica en  $\mathcal{H}(\mathbf{X})$ , haciendo de  $(\mathcal{H}(\mathbf{X}), h)$  un espacio métrico. En el plano,  $h(A, B)$  mide la “distancia” entre dos imágenes.

### 1.1.2. El papel de las transformaciones

La geometría fractal estudia tanto formas irregulares como patrones con regularidad a múltiples escalas. Muchos conjuntos pueden generarse mediante *transformaciones* simples o exhiben invariancia bajo ellas.

Una *transformación* en  $\mathbf{X}$  es una función  $f : \mathbf{X} \rightarrow \mathbf{X}$ . Para  $S \subset \mathbf{X}$ ,  $f(S) = \{f(x) : x \in S\}$ . La transformación es *inyectiva* si  $f(x) = f(y)$  implica  $x = y$ ; es *sobreyectiva* si  $f(\mathbf{X}) = \mathbf{X}$ ; es *invertible* si es biyectiva, definiéndose entonces la inversa  $f^{-1}$ . Las iteraciones hacia adelante son  $f^{\circ n}$  con  $f^{\circ 0}(x) = x$ ,  $f^{\circ(n+1)} = f \circ f^{\circ n}$ . Si  $f$  es invertible, pueden definirse iteraciones hacia atrás  $f^{\circ(-m)}$ .

Estas ideas permiten analizar cambios geométricos como contracciones, rotaciones o reflexiones, y su acción sobre conjuntos.

Una transformación  $f : \mathbf{X} \rightarrow \mathbf{X}$  es *contractiva* (o un *mapeo de contracción*) si existe  $0 \leq s < 1$  (el *factor de contracción*) tal que  $d(f(x), f(y)) \leq s \cdot d(x, y) \quad \forall x, y \in \mathbf{X}$ .

El ínfimo y supremo de un conjunto de números reales se definen de la manera usual. El teorema del punto fijo de Banach para mapeos contractivos garantiza existencia, unicidad y convergencia iterativa al punto fijo. Estos conceptos se extienden al espacio  $(\mathcal{H}(\mathbf{X}), h(d))$ .

Los mapeos contractivos son continuos. Además, un mapeo contractivo  $w$  envía  $\mathcal{H}(\mathbf{X})$  en sí mismo: si  $S$  es compacto,  $w(S)$  también lo es. En el plano,  $w(B)$  corresponde a una versión contraída de la imagen  $B$ . Definiendo  $w : \mathcal{H}(\mathbf{X}) \rightarrow \mathcal{H}(\mathbf{X})$  por  $w(B) = \{w(x) : x \in B\}$ , se obtiene un mapeo contractivo con el mismo factor  $s$ . Una propiedad útil es que para  $B, C, D, E \in \mathcal{H}(\mathbf{X})$ ,

$$h(B \cup C, D \cup E) \leq \max\{h(B, D), h(C, E)\},$$

lo que conduce al operador de Hutchinson.

**Definición 1.1.1.** Dado un espacio métrico  $(\mathbf{X}, d)$  y mapeos contractivos  $\{w_n : n = 1, \dots, N\}$  en  $\mathcal{H}(\mathbf{X})$  con factores  $s_n$ , el operador de Hutchinson  $W : \mathcal{H}(\mathbf{X}) \rightarrow \mathcal{H}(\mathbf{X})$  se define como

$$W(B) = \bigcup_{n=1}^N w_n(B) \quad \text{para } B \in \mathcal{H}(\mathbf{X}).$$

$W$  es contractivo con factor  $s = \max\{s_n\}$ .

Por ejemplo, en el plano podemos definir tres contracciones afines:

$$\begin{aligned}
w_1 \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} &= \begin{pmatrix} 0.5x_1 \\ 0.5x_2 \end{pmatrix} \\
w_2 \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} &= \begin{pmatrix} 0.5x_1 \\ 0.5x_2 + 0.5 \end{pmatrix} \\
w_3 \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} &= \begin{pmatrix} 0.5x_1 + 0.5 \\ 0.5x_2 \end{pmatrix}
\end{aligned}$$

Cada una copia, reduce y traslada una imagen inicial.

## 1.2. Sistemas de Funciones Iteradas (IFS)

Un *sistema de funciones iteradas (IFS)* es una estructura  $\{\mathbf{X}; w_n, n = 1, \dots, N\}$  donde  $(\mathbf{X}, d)$  es completo y cada  $w_n : \mathbf{X} \rightarrow \mathbf{X}$  es contractiva con factor  $s_n$ . El factor global es  $s = \max\{s_n\}$ . Un ejemplo unidimensional es  $\{\mathbb{R}; w_1, w_2\}$  con  $w_1(x) = \frac{1}{3}x$  y  $w_2(x) = \frac{1}{3}x + \frac{2}{3}$ .

Para un IFS con factor  $s$ , el operador de Hutchinson  $W(B) = \bigcup_{n=1}^N w_n(B)$  es contractivo en  $(\mathcal{H}(\mathbf{X}), h(d))$  con el mismo  $s$ . Su único punto fijo  $A \in \mathcal{H}(\mathbf{X})$ , llamado *atractor* del IFS, satisface

$$A = W(A) = \bigcup_{n=1}^N w_n(A)$$

y se obtiene como  $A = \lim_{n \rightarrow \infty} W^{on}(B)$  para cualquier  $B \in \mathcal{H}(\mathbf{X})$ .

Consideremos un IFS sobre el cuadrado unitario  $\blacksquare = [0, 1] \times [0, 1]$  con las tres contracciones afines antes definidas. Partiendo de una imagen inicial (Figura 1.1), la aplicación iterada de  $W$  produce una secuencia de imágenes que, tras varias iteraciones, se estabiliza en el atractor, un fractal conocido como variante del triángulo de Sierpinski.

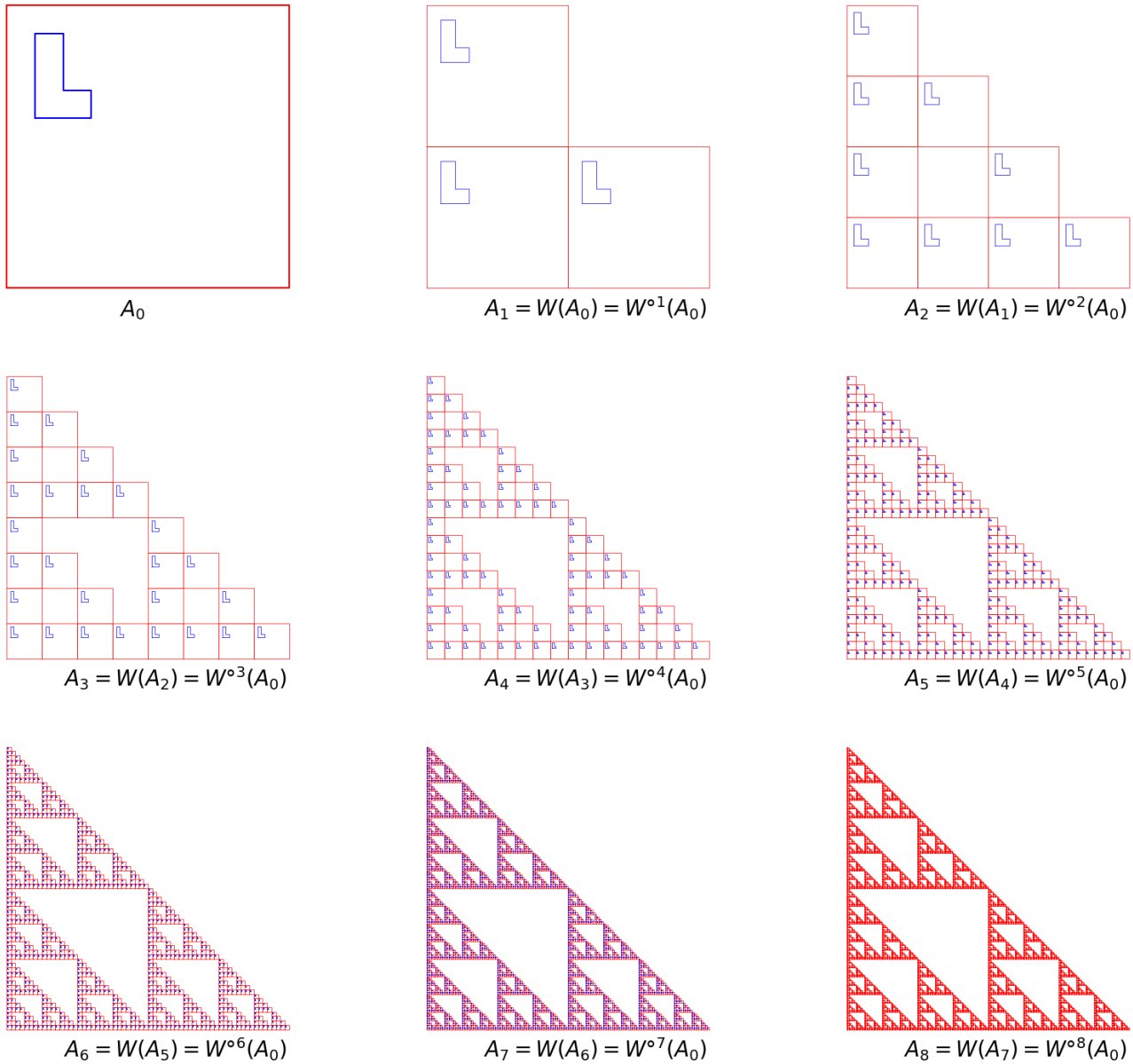
### 1.2.1. El triángulo de Sierpinski

La Figura 1.1 muestra una variante del conocido triángulo de Sierpinski, atractor de un IFS con tres transformaciones.

Las transformaciones afines en el plano pueden escribirse en forma matricial:

$$w_i \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} a_i & b_i \\ c_i & d_i \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} + \begin{pmatrix} e_i \\ f_i \end{pmatrix}.$$

Los parámetros  $(a_i, b_i, c_i, d_i, e_i, f_i)$  caracterizan completamente cada transformación. Para el triángulo de Sierpinski estándar, estos parámetros son como en la Tabla 1.1.



**Figura 1.1:** Una variante del triángulo de Sierpinski generada por un IFS de tres transformaciones afines con factor  $s = 1/2$ . Se observan las primeras ocho iteraciones a partir de  $A_0$ .

**Tabla 1.1:** Parámetros del IFS para el triángulo de Sierpinski

| $w$ | $a$ | $b$ | $c$ | $d$ | $e$ | $f$ |
|-----|-----|-----|-----|-----|-----|-----|
| 1   | 0.5 | 0   | 0   | 0.5 | 0   | 0   |
| 2   | 0.5 | 0   | 0   | 0.5 | 0.5 | 0   |
| 3   | 0.5 | 0   | 0   | 0.5 | 0   | 0.5 |

El *algoritmo de iteraciones aleatorias* para un IFS asigna probabilidades  $p_i > 0$  (con  $\sum p_i = 1$ ) a cada transformación. Partiendo de un punto  $x_0$ , se elige iterativamente  $x_n \in \{w_1(x_{n-1}), \dots, w_N(x_{n-1})\}$  según esas probabilidades, generando una trayectoria  $\{x_n\}$  que aproxima el atractor.

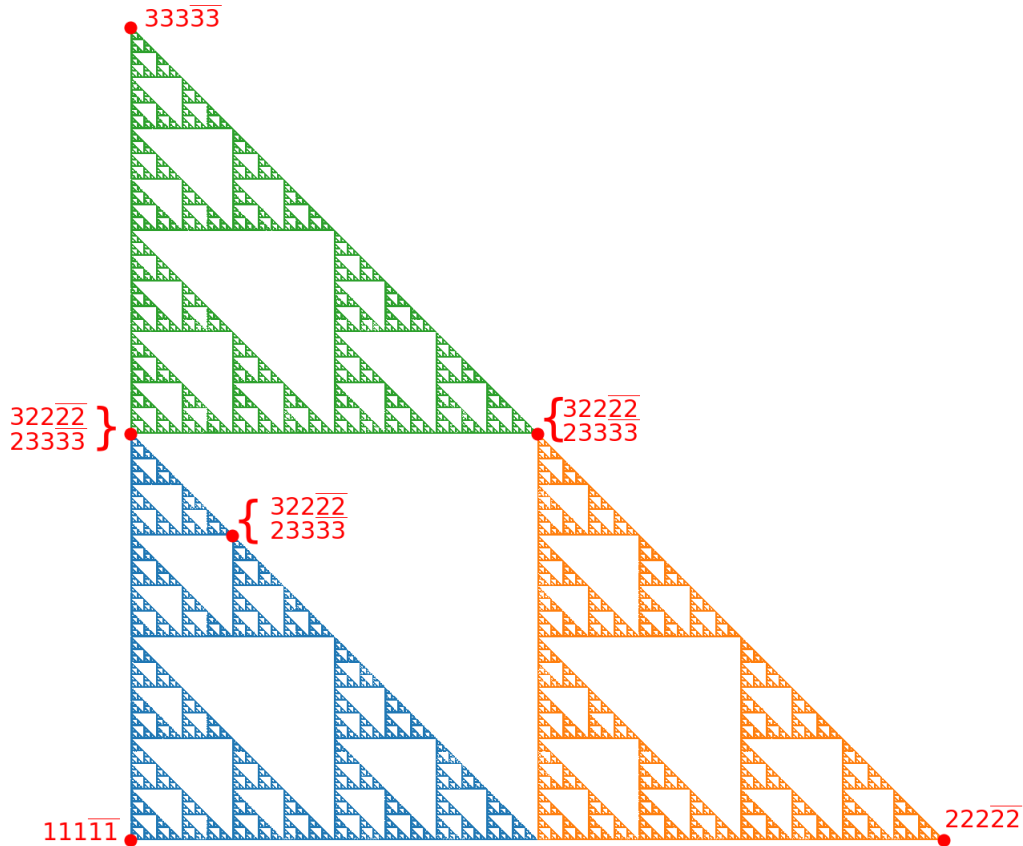
Este proceso recuerda fenómenos como el *movimiento browniano*, donde una partícula se desplaza aleatoriamente en un medio.

### 1.3. Direcciones en un fractal

Consideremos el IFS en  $\mathbb{C}$ :

$$\{\mathbb{C}; w_1(z) = \frac{1}{2}z, w_2(z) = \frac{1}{2}z + \frac{1}{2}, w_3(z) = \frac{1}{2}z + \frac{1}{2}i\}.$$

Su atractor es el triángulo de Sierpiński con vértices  $(0,0)$ ,  $(1,0)$  y  $(0,1)$ . A cada punto de  $A$  puede asignársele una “dirección” (secuencia infinita de índices) según la historia de transformaciones que lo generan. Algunos puntos tienen múltiples direcciones debido a intersecciones no vacías entre  $w_i(A)$  y  $w_j(A)$  (Figura 1.2). Otros, como los vértices, tienen dirección única. Cuando estas intersecciones son no vacías pero “pequeñas” (en sentido técnico), el IFS se denomina *solo tocando*. Si las intersecciones son significativas, el IFS es *superpuesto*.



**Figura 1.2:** En el triángulo de Sierpiński, ciertos puntos poseen dos direcciones, otros solo una.

El espacio de direcciones  $\Sigma$  sobre  $N$  símbolos (enteros  $\{0, \dots, N-1\}$ ) consiste en secuencias infinitas de símbolos. Un conjunto es *contable* si puede ponerse en correspondencia sobreyectiva con

un subconjunto de los enteros positivos; de lo contrario, es *incontable*.  $\Sigma$  con dos o más símbolos es incontable.

## 1.4. Nociones de sistemas dinámicos

Un *sistema dinámico* es un par  $\{\mathbf{X}, f\}$  donde  $f : \mathbf{X} \rightarrow \mathbf{X}$  es una transformación en un espacio métrico. La *órbita* de  $x$  es  $\{f^{on}(x)\}_{n=0}^{\infty}$ . Un punto  $x$  es *periódico* si  $f^{on}(x) = x$  para algún  $n \geq 1$ ; el menor tal  $n$  es su *periodo mínimo*. La órbita de un punto periódico es un *ciclo*. Un punto fijo  $x_f$  es *atractivo* (estable) si  $f$  es contractiva en un entorno de  $x_f$ ; es *repulsivo* (inestable) si, en un entorno,  $f$  expande distancias. Un punto es *eventualmente periódico* si alguna iterada suya es periódica.

Se dice que  $f$  exhibe *mezcla* si, para cualesquiera conjuntos abiertos  $A$  y  $B$ , existe  $N$  tal que  $f^{on}(A) \cap B \neq \emptyset$  para  $n > N$ . Esto implica la existencia de órbitas densas. Un sistema dinámico es *transitivo* si para abiertos  $\mathcal{U}, \mathcal{V}$  existe  $n$  con  $\mathcal{U} \cap f^{on}(\mathcal{V}) \neq \emptyset$ . Es *sensible a condiciones iniciales* si pequeñas diferencias iniciales se amplían. En este trabajo vamos a usar la definición matemática de sistema *caótico* como uno que es transitivo, sensible y tiene órbitas periódicas densas.

## 1.5. Dinámica sobre fractales

Para un IFS  $\{\mathbf{X}; w_n\}$  con atractor  $A$ , si el IFS es totalmente desconectado, cada  $w_n : A \rightarrow A$  es inyectiva. Definiendo  $M = \bigcup_{i \neq j} (w_i(A) \cap w_j(A))$ , si  $M$  contiene un abierto relativo, el IFS está superpuesto; si  $M$  tiene interior vacío y las transformaciones son afines, está solo tocando.

El sistema dinámico de desplazamiento  $\{A; S\}$  asociado a un IFS totalmente desconectado es topológicamente conjugado al sistema de desplazamiento  $\{\Sigma; T\}$  en el espacio de direcciones, mediante el homeomorfismo  $\phi : \Sigma \rightarrow A$ . Esto implica correspondencia entre ciclos repulsivos en  $A$  y en  $\Sigma$ .

## 1.6. La noción de dimensión fractal

Para comparar fractales se utilizan números que cuantifican su complejidad, usualmente relacionados con su *dimensión*.

Dado  $A \in \mathcal{H}(X)$ , sea  $\mathcal{N}(A, \epsilon)$  el mínimo número de bolas cerradas de radio  $\epsilon$  necesarias para cubrir  $A$ . Si existe

$$D = \lim_{\epsilon \rightarrow 0} \frac{\ln \mathcal{N}(A, \epsilon)}{\ln(1/\epsilon)},$$

se llama a  $D$  la *dimensión fractal* de  $A$ , denotada  $D(A)$ .

Por ejemplo, la dimensión de un punto es 0. Para simplificar, si  $\epsilon_n = Cr^n$  con  $0 < r < 1$ ,  $C > 0$ ,



y existe

$$D = \lim_{n \rightarrow \infty} \frac{\ln \mathcal{N}(A, \epsilon_n)}{\ln(1/\epsilon_n)},$$

entonces  $A$  tiene dimensión  $D$ . En  $\mathbb{R}^m$ , usando cajas de lado  $1/2^n$ , con  $\mathcal{N}_n(A)$  el número de cajas que intersectan  $A$ , tenemos

$$D = \lim_{n \rightarrow \infty} \frac{\ln \mathcal{N}_n(A)}{\ln(2^n)}.$$

Para el triángulo de Sierpinski,  $\mathcal{N}_n = 3^n$ , luego  $D = \ln 3 / \ln 2$ .

Alternativamente, puede definirse  $D$  mediante un límite superior. En  $(\mathbb{R}^m, \text{Euclidiana})$ , la dimensión fractal existe para todo conjunto en  $\mathcal{H}(\mathbb{R}^m)$  y satisface  $0 \leq D(A) \leq m$ . Además, si  $A \subset B$ , entonces  $D(A) \leq D(B)$ .

Para un IFS  $\{\mathbb{R}^m; w_1, \dots, w_N\}$  con atractor  $A$ , donde cada  $w_n$  es afín con factor de escala  $s_n$ , si el IFS está totalmente desconectado o solo tocándose,  $D(A)$  es la única solución de

$$\sum_{n=1}^N |s_n|^D = 1, \quad D \in [0, m].$$

Si el IFS está superpuesto, entonces  $\overline{D} \geq D(A)$ , donde  $\overline{D}$  resuelve la misma ecuación (permitiendo  $\overline{D} \in [0, \infty)$ ).

Para el conjunto de Cantor generado por dos contracciones de factor  $r$ , la ecuación es  $2r^D = 1$ , luego  $D = \log 2 / \log(1/r)$ .

La *dimensión de Hausdorff-Besicovitch*  $D_H(A)$  es un concepto más refinado. Para un conjunto acotado  $A \subset \mathbb{R}^m$ , se define el diámetro  $\text{diam}(A) = \sup\{d(x, y) : x, y \in A\}$ . Dado  $p \geq 0$  y  $\epsilon > 0$ , se considera

$$\mathcal{M}(A, p, \epsilon) = \inf \left\{ \sum_i (\text{diam}(A_i))^p : A \subset \bigcup_i A_i, \text{diam}(A_i) < \epsilon \right\}.$$

Luego  $\mathcal{M}(A, p) = \sup_{\epsilon > 0} \mathcal{M}(A, p, \epsilon)$  es la *medida  $p$ -dimensional de Hausdorff*. Existe un único  $D_H \in [0, m]$  tal que  $\mathcal{M}(A, p) = \infty$  si  $p < D_H$  y  $\mathcal{M}(A, p) = 0$  si  $p > D_H$ . Este  $D_H$  es la dimensión de Hausdorff-Besicovitch. Se cumple  $0 \leq D_H(A) \leq D(A) \leq m$ .

En este capítulo se han introducido la dimensión fractal y la de Hausdorff-Besicovitch, conceptos útiles para distinguir fractales, aunque a veces insuficientes, requiriéndose entonces otras cantidades para una caracterización completa.



## Capítulo 2

# Fundamentos de las Medidas Invariantes en Estructuras Fractales

El presente capítulo aborda los principios teóricos esenciales concernientes a las medidas que se mantienen invariantes en el ámbito de los conjuntos fractales, partiendo de una extensión de los sistemas de funciones iteradas (IFS) mediante la asignación de pesos probabilísticos. Se introduce y analiza el operador de Markov, así como un resultado fundamental conocido como teorema de Elton, el cual establece un vínculo preciso entre las órbitas de carácter aleatorio y la medida invariante. La exposición prosigue con la generalización de estos conceptos a los IFS recurrentes, examinando sus propiedades distintivas y su relación con las cadenas de Markov. Finalmente, se presenta el teorema de Collage en el contexto de IFS recurrentes y se sientan las bases del estudio de los vectores de medida, proporcionando así el marco necesario para abordar configuraciones fractales de naturaleza más intrincada.

De manera formal, un sistema de funciones iteradas dotado de probabilidades se compone de un IFS estándar

$$\{\mathbf{X}; f_1, f_2, \dots, f_N\}$$

al cual se asocia un vector de probabilidades  $\{q_1, q_2, \dots, q_N\}$ . Este vector debe satisfacer las condiciones de normalización y positividad:

$$\sum_{j=1}^N q_j = 1 \quad \text{y} \quad q_j > 0 \quad \text{para todo } j = 1, 2, \dots, N.$$

Cada coeficiente  $q_j$  se asocia de manera biunívoca con la transformación contractiva correspondiente  $f_j$ . La notación completa que describe este objeto es

$$\{\mathbf{X}; f_1, f_2, \dots, f_N; q_1, q_2, \dots, q_N\},$$

siendo aceptable omitir la referencia explícita a las probabilidades cuando el contexto no induzca confusión.

Desde una perspectiva aplicada, por ejemplo en la generación de imágenes en tonos de gris, los parámetros  $q_j$  permiten modular la densidad con la que los puntos se distribuyen sobre la imagen final. Al calcular la frecuencia relativa de la ocurrencia de puntos en subregiones específicas del atractor, como las imágenes  $f_j(\blacksquare)$ , es posible inducir una medida bien definida sobre dicho atractor. Esta medida posee la propiedad de invariancia y encuentra una interpretación natural como la intensidad del tono gris en la correspondiente porción de la imagen.

Para ilustrar estos conceptos, considérese el siguiente IFS con asignación probabilística:

$$\{\mathbb{C}; f_1(z), f_2(z), f_3(z), f_4(z); 0.1, 0.2, 0.3, 0.4\},$$

definido mediante las transformaciones afines

$$\begin{aligned} f_1(z) &= 0.5z, \\ f_2(z) &= 0.5z + 0.5, \\ f_3(z) &= 0.5z + 0.5i, \\ f_4(z) &= 0.5z + 0.5 + 0.5i. \end{aligned}$$

Los parámetros completos de este sistema, incluidos los coeficientes de las transformaciones y sus probabilidades asociadas, se sintetizan en la Tabla 2.1.

**Tabla 2.1:** Coeficientes de las transformaciones y probabilidades  $q$  para un IFS.

| $f$ | $a$ | $b$ | $c$ | $d$ | $e$ | $f$ | $q$ |
|-----|-----|-----|-----|-----|-----|-----|-----|
| 1   | 0.5 | 0   | 0   | 0.5 | 0   | 0   | 0.1 |
| 2   | 0.5 | 0   | 0   | 0.5 | 0.5 | 0   | 0.2 |
| 3   | 0.5 | 0   | 0   | 0.5 | 0   | 0.5 | 0.3 |
| 4   | 0.5 | 0   | 0   | 0.5 | 0.5 | 0.5 | 0.4 |

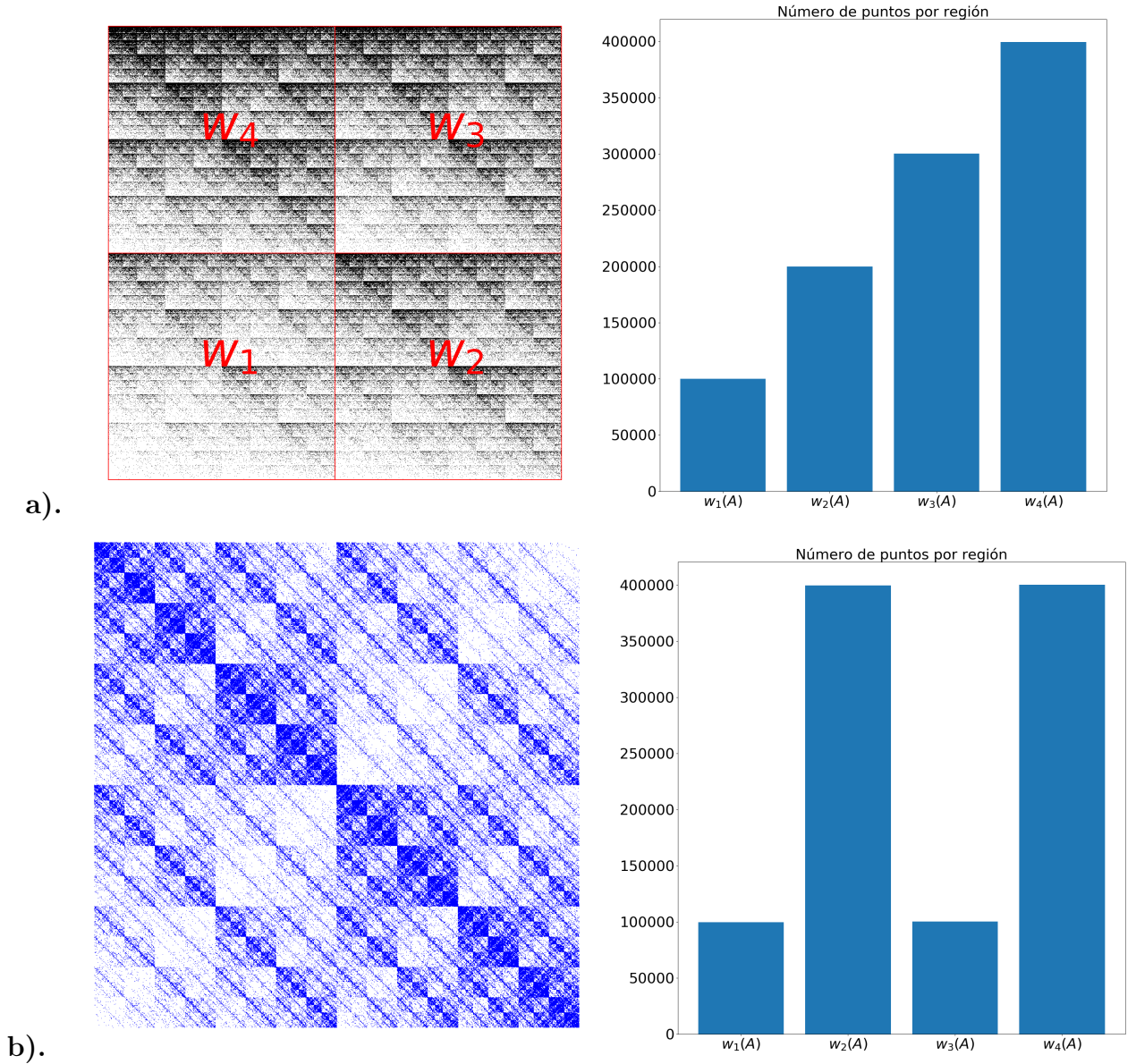
El objeto geométrico límite, o atractor, para este conjunto de transformaciones es el cuadrado unitario  $\blacksquare$ . El método práctico para su aproximación es el algoritmo de iteración aleatoria, cuyo procedimiento se detalla a continuación:

Se inicia el proceso seleccionando un punto arbitrario  $x_0$  perteneciente a  $\mathbb{C}$ . En cada paso discreto del algoritmo, se elige una transformación  $f_j$  del conjunto disponible, donde la probabilidad de seleccionar  $f_j$  es precisamente  $q_j$  para  $j = 1, 2, 3, 4$ . Esta transformación se aplica al punto corriente, generando así un nuevo punto. Suponiendo, por conveniencia, que  $x_0 \in \blacksquare$ , y dado que  $f_j(\blacksquare) \subseteq \blacksquare$  para toda  $j$ , la secuencia completa de puntos  $\{x_n : n = 1, 2, \dots, M\}$  (con  $M$  denotando el número total de iteraciones) permanecerá contenida dentro del cuadrado unitario.

Al ejecutar este algoritmo con los parámetros de la Tabla 2.1 y un número suficientemente elevado de iteraciones, se obtiene una nube de puntos que aproxima visualmente al atractor  $\blacksquare$ . La densidad

de esta nube en diferentes subregiones no es uniforme, sino que refleja las probabilidades asignadas. Por ejemplo, para  $M = 10,000$ , la cantidad esperada de puntos en la región  $f_1(\blacksquare)$  es cercana a 1,000, en  $f_2(\blacksquare)$  es de aproximadamente 2,000, en  $f_3(\blacksquare)$  ronda los 3,000, y en  $f_4(\blacksquare)$  se acerca a 4,000.

La Figura 2.1 proporciona una verificación empírica de este comportamiento, mostrando los resultados de simular el algoritmo con  $M = 1,000,000$ .



**Figura 2.1:** **a).** Representación gráfica obtenida mediante el algoritmo de iteración aleatoria aplicado al sistema de la Tabla 2.1, acompañada de un histograma que cuantifica la ocupación en cada región  $f_j(A)$ . **b).** Resultado análogo para el mismo sistema de transformaciones pero con un vector de probabilidades distinto. Se aprecia una textura visual diferente a pesar de que el soporte geométrico (el atractor) es idéntico en ambos casos.

El panel b) de la Figura 2.1 muestra cómo, al modificar las probabilidades asignadas a las mismas transformaciones, se altera la textura de la imagen generada. El atractor subyacente como conjunto de puntos permanece siendo  $\blacksquare$ , pero la distribución de la masa (o densidad de puntos) sobre él varía significativamente.

En un marco teórico más general, sea  $(\mathbf{X}, d)$  un espacio métrico completo y  $\{\mathbf{X}; f_1, \dots, f_N; q_1, \dots, q_N\}$

un IFS con probabilidades. Si denotamos por  $A$  su atractor, existe una única medida de Borel regular  $\nu$ , denominada **medida invariante**, que satisface propiedades particulares. Dicha medida asigna un valor no negativo, interpretable como “masa” o “peso”, a subconjuntos adecuados de  $\mathbf{X}$ .

Esta medida cumple con  $\nu(A) = 1$ , indicando que la totalidad de la masa se concentra en el atractor, mientras que  $\nu(\emptyset) = 0$ . Adicionalmente, se tiene que  $\nu(\mathbf{X}) = 1$ . No todos los subconjuntos de  $\mathbf{X}$  son medibles en el sentido requerido; la familia de subconjuntos para los cuales  $\nu$  está definida son los *conjuntos de Borel* de  $\mathbf{X}$ , denotados por  $\mathcal{B}(\mathbf{X})$ . Toda región abierta  $\mathcal{O} \subset \mathbf{X}$  pertenece a esta familia.

La conexión profunda con el algoritmo de iteración aleatoria se manifiesta en el siguiente hecho: dada la sucesión  $\{x_n\}_{n=0}^{\infty}$  generada por el algoritmo, y definiendo  $\mathcal{N}(B, n)$  como el cardinal del conjunto  $\{x_0, x_1, \dots, x_n\} \cap B$ , se verifica que, con probabilidad uno,

$$\nu(B) = \lim_{n \rightarrow \infty} \frac{\mathcal{N}(B, n)}{n + 1}$$

para conjuntos  $B \in \mathcal{B}(\mathbf{X})$  cuya frontera tenga medida nula ( $\nu(\partial B) = 0$ ). En esencia, la medida invariante  $\nu$  de un conjunto coincide con la frecuencia asintótica con la que la órbita aleatoria visita dicho conjunto.

Este principio ofrece un método empírico para definir distribuciones de probabilidad en sistemas dinámicos más generales. El cociente  $\frac{\mathcal{N}(B, n)}{n+1}$  constituye una frecuencia relativa empírica, cuya convergencia define la medida teórica  $\nu$ , la cual funciona como una distribución de probabilidad sobre el espacio de fases. Existe una analogía instructiva con la mecánica estadística de equilibrio: los macroestados de un sistema termodinámico se describen mediante distribuciones de probabilidad sobre el espacio de microestados (el espacio de fases en la mecánica clásica, o el espacio de Hilbert en la cuántica). De modo similar a la subdivisión del atractor en la Figura 2.1, el espacio de fases puede ser particionado en celdas, permitiendo calcular frecuencias relativas que conducen a la definición de probabilidades macroscópicas.

## 2.1. Álgebras y $\sigma$ -Álgebras de Conjuntos

Ofrecemos aquí una síntesis de los elementos básicos de la teoría de la medida, disciplina íntimamente ligada a la teoría de probabilidades. Si bien sus formulaciones pueden alcanzar un alto grado de abstracción, sus orígenes y motivaciones se encuentran en cuestiones concretas y aplicadas, como las exploradas en la sección precedente.

**Definición 2.1.1.** *Dado un conjunto  $\mathbf{X}$ , una colección no vacía  $\mathcal{A}$  compuesta por subconjuntos de  $\mathbf{X}$  recibe el nombre de álgebra de conjuntos (o campo) cuando verifica las dos condiciones siguientes:*

1. Si  $A, B$  son elementos de  $\mathcal{A}$ , su reunión  $A \cup B$  también pertenece a  $\mathcal{A}$ ;

2. Si  $A$  es un elemento de  $\mathcal{A}$ , su complemento relativo  $\mathbf{X} \setminus A$  también es un elemento de  $\mathcal{A}$ .

De lo anterior se deduce que el conjunto total  $\mathbf{X}$  es miembro de toda álgebra  $\mathcal{A}$ . En efecto, si tomamos cualquier  $A \in \mathcal{A}$ , su complemento  $\mathbf{X} \setminus A$  está en  $\mathcal{A}$ , y por tanto la unión  $A \cup (\mathbf{X} \setminus A) = \mathbf{X}$  pertenece a  $\mathcal{A}$ . Consecuentemente, el conjunto vacío  $\emptyset = \mathbf{X} \setminus \mathbf{X}$  también forma parte de la álgebra.

**Teorema 2.1.1.** *Sea  $\mathbf{X}$  un conjunto y sea  $\mathcal{G}$  una familia no vacía de subconjuntos de  $\mathbf{X}$ . Consideremos la familia  $\mathcal{A}$  que incluye a todos los subconjuntos de  $\mathbf{X}$  que es posible obtener a partir de miembros de  $\mathcal{G}$  mediante la aplicación sucesiva de las operaciones de unión, intersección y complementación con respecto a  $\mathbf{X}$ . Esta familia  $\mathcal{A}$  constituye una álgebra de conjuntos.*

Una demostración detallada de este resultado puede hallarse en el capítulo nueve de la referencia [2].

**Definición 2.1.2.** *La álgebra  $\mathcal{A}$  construida en el Teorema 2.1.1 recibe el nombre de álgebra generada por la familia  $\mathcal{G}$ .*

Como ilustración, si tomamos un subconjunto  $A$  de un conjunto  $\mathbf{X}$ , la colección  $\mathcal{A} = \{\mathbf{X}, A, \mathbf{X} \setminus A, \emptyset\}$  satisface las propiedades de una álgebra.

**Definición 2.1.3.** *Una álgebra de conjuntos  $\mathcal{A}$  se denomina  $\sigma$ -álgebra (sigma-álgebra) cuando es cerrada bajo la operación de unión numerable; es decir, si satisface la condición adicional:*

$$\text{Si } A_i \in \mathcal{A} \text{ para } i = 1, 2, 3, \dots, \text{ entonces } \bigcup_{i=1}^{\infty} A_i \in \mathcal{A}.$$

Dada cualquier álgebra, es posible considerar la  $\sigma$ -álgebra más pequeña que la contiene.

**Teorema 2.1.2.** *Sea  $\mathbf{X}$  un conjunto y sea  $\mathcal{G}$  una familia de subconjuntos de  $\mathbf{X}$ . Sea  $\{\mathcal{F}_\alpha : \alpha \in I\}$  la colección de todas las  $\sigma$ -álgebras definidas sobre  $\mathbf{X}$  que contienen a  $\mathcal{G}$ . Entonces, la intersección  $\mathcal{F} = \bigcap_{\alpha \in I} \mathcal{F}_\alpha$  es una  $\sigma$ -álgebra.*

**Demostración:** Es necesario probar dos hechos: que  $\mathcal{F}$  es una  $\sigma$ -álgebra y que cada  $\mathcal{F}_\alpha$  contiene a  $\mathcal{G}$  por definición.

Supongamos que  $\{B_i\}_{i=1}^{\infty}$  es una sucesión de conjuntos tal que cada  $B_i$  pertenece a  $\mathcal{F}$ . Esto significa que, para cada índice  $\alpha$ , se tiene  $B_i \in \mathcal{F}_\alpha$ . Como cada  $\mathcal{F}_\alpha$  es una  $\sigma$ -álgebra, la unión  $\bigcup_{i=1}^{\infty} B_i$  es un elemento de  $\mathcal{F}_\alpha$ . Por consiguiente, esta unión también pertenece a la intersección  $\mathcal{F}$ .

Ahora, si  $B$  es un conjunto en  $\mathcal{F}$ , entonces  $B \in \mathcal{F}_\alpha$  para todo  $\alpha$ . La propiedad de cerradura bajo complementación de cada  $\sigma$ -álgebra  $\mathcal{F}_\alpha$  implica que  $\mathbf{X} \setminus B \in \mathcal{F}_\alpha$  para todo  $\alpha$ . En consecuencia,  $\mathbf{X} \setminus B$  es un elemento de  $\mathcal{F}$ .

Esto completa la verificación de que  $\mathcal{F}$  es una  $\sigma$ -álgebra.

**Definición 2.1.4.** Dada una familia  $\mathcal{G}$  de subconjuntos de  $\mathbf{X}$ , la  $\sigma$ -álgebra  $\mathcal{F}$  definida en el Teorema 2.1.2 se denomina  $\sigma$ -álgebra generada por  $\mathcal{G}$ .

**Definición 2.1.5.** Sea  $(\mathbf{X}, d)$  un espacio métrico. La  $\sigma$ -álgebra de Borel  $\mathcal{B}$  asociada a este espacio es la  $\sigma$ -álgebra generada por la colección de todos los subconjuntos abiertos de  $\mathbf{X}$ . Un elemento cualquiera de  $\mathcal{B}$  se denomina conjunto de Borel de  $\mathbf{X}$ .

El teorema que sigue justifica una forma constructiva de obtener la  $\sigma$ -álgebra de Borel en contextos con suficiente estructura.

**Teorema 2.1.3.** Si  $(\mathbf{X}, d)$  es un espacio métrico compacto, entonces la  $\sigma$ -álgebra de Borel asociada puede ser generada por una familia numerable de bolas abiertas. (Para una demostración, véase [2]).

## 2.2. El Concepto de Medida

La noción de medida se introduce sobre una álgebra (o  $\sigma$ -álgebra) de conjuntos. Consiste en asignar a cada elemento de dicha álgebra un número real no negativo, interpretado como su “tamaño”, “masa” o “probabilidad”. En el desarrollo formal se distingue entre medidas definidas sobre álgebras (a veces llamadas pre-medidas) y su extensión a  $\sigma$ -álgebras, proceso garantizado por el importante teorema de extensión de Carathéodory.

**Definición 2.2.1.** Una medida  $\mu$  sobre una álgebra  $\mathcal{A}$  es una función  $\mu : \mathcal{A} \rightarrow [0, +\infty]$  que cumple con la condición de  $\sigma$ -aditividad: para toda sucesión  $\{A_i\}_{i=1}^{\infty}$  de conjuntos disjuntos dos a dos (esto es,  $A_i \cap A_j = \emptyset$  cuando  $i \neq j$ ) pertenecientes a  $\mathcal{A}$  y tales que su unión  $\bigcup_{i=1}^{\infty} A_i$  también pertenece a  $\mathcal{A}$ , se satisface

$$\mu \left( \bigcup_{i=1}^{\infty} A_i \right) = \sum_{i=1}^{\infty} \mu(A_i).$$

**Definición 2.2.2.** Sea  $(\mathbf{X}, d)$  un espacio métrico y sea  $\mathcal{B}$  su  $\sigma$ -álgebra de Borel. Una medida  $\mu$  definida sobre  $\mathcal{B}$  recibe el nombre de medida de Borel.

**Teorema 2.2.1.** Sea  $\mu$  una medida definida sobre una álgebra  $\mathcal{A}$ . Entonces  $\mu$  posee las siguientes propiedades:

1. (Aditividad finita) Si  $A, B \in \mathcal{A}$  son disjuntos, entonces  $\mu(A \cup B) = \mu(A) + \mu(B)$ .
2. (Monotonía) Si  $A, B \in \mathcal{A}$  con  $A \subset B$ , entonces  $\mu(A) \leq \mu(B)$ .
3.  $\mu(\emptyset) = 0$ .
4. (Subaditividad numerable) Si  $\{A_i\}_{i=1}^{\infty} \subset \mathcal{A}$  y  $\bigcup_{i=1}^{\infty} A_i \in \mathcal{A}$ , entonces  $\mu(\bigcup_{i=1}^{\infty} A_i) \leq \sum_{i=1}^{\infty} \mu(A_i)$ .



5. (Continuidad desde abajo) Si  $\{A_i\}_{i=1}^{\infty} \subset \mathcal{A}$  es una sucesión creciente ( $A_1 \subset A_2 \subset \dots$ ) con  $\bigcup_{i=1}^{\infty} A_i \in \mathcal{A}$ , entonces  $\lim_{i \rightarrow \infty} \mu(A_i) = \mu(\bigcup_{i=1}^{\infty} A_i)$ .
6. (Continuidad desde arriba) Si  $\{A_i\}_{i=1}^{\infty} \subset \mathcal{A}$  es una sucesión decreciente ( $A_1 \supset A_2 \supset \dots$ ) con  $\bigcap_{i=1}^{\infty} A_i \in \mathcal{A}$  y  $\mu(A_1) < \infty$ , entonces  $\lim_{i \rightarrow \infty} \mu(A_i) = \mu(\bigcap_{i=1}^{\infty} A_i)$ .

Las demostraciones de estas propiedades se pueden consultar en [27].

En el análisis de espacios métricos, como  $\mathbb{R}^n$  con la distancia euclidiana, las medidas de Borel juegan un papel central. El teorema fundamental de Carathéodory asegura que una medida definida sobre una álgebra que genere la  $\sigma$ -álgebra de Borel puede extenderse de manera única a toda esta  $\sigma$ -álgebra.

**Teorema 2.2.2** (Teorema de Extensión de Carathéodory). Sea  $\mu$  una medida definida sobre una álgebra  $\mathcal{A}$  de subconjuntos de un conjunto  $X$ . Sea  $\widehat{\mathcal{A}}$  la  $\sigma$ -álgebra generada por  $\mathcal{A}$ . Entonces existe una única medida  $\widehat{\mu}$  definida sobre  $\widehat{\mathcal{A}}$  que extiende a  $\mu$ ; es decir, tal que  $\widehat{\mu}(A) = \mu(A)$  para todo  $A \in \mathcal{A}$ .

**Esquema de la demostración:** Se define la medida exterior  $\mu^*$  asociada a  $\mu$  para cualquier subconjunto  $E \subset X$  mediante:

$$\mu^*(E) = \inf \left\{ \sum_{n=1}^{\infty} \mu(A_n) : E \subset \bigcup_{n=1}^{\infty} A_n, A_n \in \mathcal{A} \right\}.$$

Un conjunto  $M \subset X$  se declara  $\mu^*$ -medible si para todo conjunto  $A \in \mathcal{A}$  se verifica la igualdad

$$\mu^*(A) = \mu^*(M \cap A) + \mu^*(A \setminus M).$$

La colección  $\mathcal{M}$  de todos los conjuntos  $\mu^*$ -medibles constituye una  $\sigma$ -álgebra que contiene a  $\mathcal{A}$  (y por tanto a  $\widehat{\mathcal{A}}$ ), y la restricción de  $\mu^*$  a  $\mathcal{M}$  es una medida que extiende a  $\mu$ . La unicidad se garantiza bajo condiciones de finitud apropiadas. Para los detalles completos, véase [27], páginas 67-70.

**Teorema 2.2.3.** Sea  $\mu$  una medida sobre una álgebra  $\mathcal{A}$ , y sea  $\widehat{\mu}$  su extensión a la  $\sigma$ -álgebra  $\widehat{\mathcal{A}}$  generada por  $\mathcal{A}$ . Entonces, para todo conjunto  $B \in \widehat{\mathcal{A}}$ , el valor  $\widehat{\mu}(B)$  puede obtenerse mediante la fórmula:

$$\widehat{\mu}(B) = \inf \left\{ \sum_{n=1}^{\infty} \mu(A_n) : B \subset \bigcup_{n=1}^{\infty} A_n, A_n \in \mathcal{A} \forall n \right\}.$$

**Definición 2.2.3.** Sea  $(\mathbf{X}, d)$  un espacio métrico y sea  $\mu$  una medida de Borel definida sobre él. Se define el soporte de  $\mu$ , denotado por  $\text{sop}(\mu)$ , como el conjunto de todos los puntos  $x \in \mathbf{X}$  para los cuales toda bola abierta centrada en  $x$  tiene medida positiva:

$$\text{sop}(\mu) = \{x \in \mathbf{X} : \mu(B(x, \epsilon)) > 0 \text{ para todo } \epsilon > 0\},$$

donde  $B(x, \epsilon) = \{y \in \mathbf{X} : d(y, x) < \epsilon\}$ .

Intuitivamente, el soporte de una medida es el subconjunto cerrado más pequeño fuera del cual la medida se anula.

**Teorema 2.2.4.** *El soporte de una medida de Borel  $\mu$  en un espacio métrico  $(\mathbf{X}, d)$  es siempre un subconjunto cerrado de  $\mathbf{X}$ .*

## 2.3. Integrales y Métricas para Medidas

En este apartado se desarrolla la teoría de integración con respecto a una medida, con el propósito fundamental de definir una distancia en conjuntos cuyos elementos son medidas. Bajo hipótesis adecuadas, el espacio métrico resultante es compacto.

Supongamos que  $(\mathbf{X}, d)$  es un espacio métrico compacto y que  $\mu$  es una medida de Borel definida sobre  $\mathbf{X}$ . Dada una función continua  $f : \mathbf{X} \rightarrow \mathbb{R}$ , deseamos asignar un significado preciso a expresiones del tipo

$$\int_{\mathbf{X}} f(x) d\mu(x).$$

**Definición 2.3.1.** *Para un subconjunto  $A \subseteq \mathbf{X}$ , la función característica (o indicatriz) de  $A$  se denota por  $\mathbf{1}_A$  y se define como*

$$\mathbf{1}_A(x) = \begin{cases} 1, & x \in A, \\ 0, & x \notin A. \end{cases}$$

*Una función  $f : \mathbf{X} \rightarrow \mathbb{R}$  se dice simple si puede escribirse como combinación lineal finita de funciones características de conjuntos de Borel; es decir, si existen números reales  $a_1, a_2, \dots, a_k$  y conjuntos de Borel disjuntos  $B_1, B_2, \dots, B_k$  tales que*

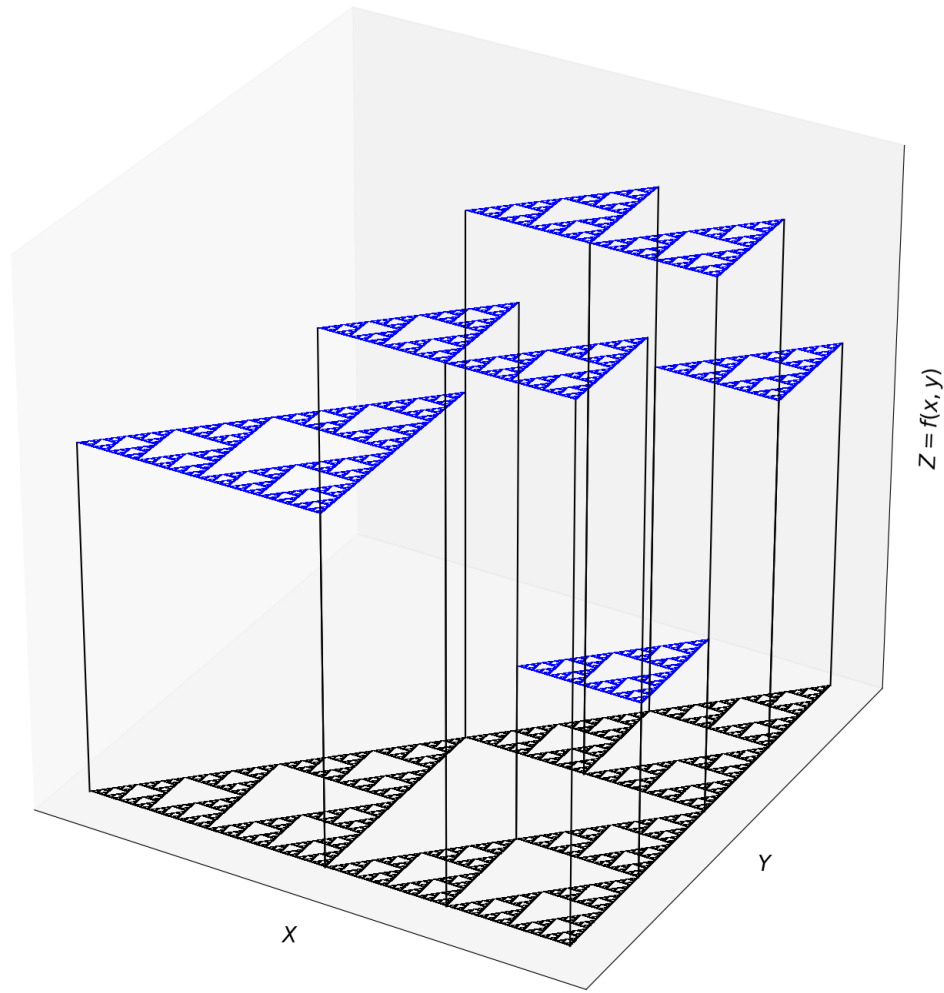
$$f(x) = \sum_{j=1}^k a_j \mathbf{1}_{B_j}(x),$$

*y además  $\bigcup_{j=1}^k B_j = \mathbf{X}$ .*

Toda función que tome únicamente un número finito de valores reales es, por tanto, una función simple. En la Figura 2.2 se ilustra un ejemplo de tal función, definida sobre el atractor fractal conocido como triángulo de Sierpinski  $\triangle$ , donde la variable dependiente representa la altura  $f(X, Y)$ .

**Definición 2.3.2.** *Para una función simple  $f(x) = \sum_{j=1}^k a_j \mathbf{1}_{B_j}(x)$ , se define su integral con respecto a la medida  $\mu$  mediante la fórmula*

$$\int_{\mathbf{X}} f(x) d\mu(x) = \sum_{j=1}^k a_j \mu(B_j).$$



**Figura 2.2:** Ejemplo de una función simple cuyo dominio es el triángulo de Sierpinski.

Cuando  $\mu$  es la medida de Lebesgue sobre el intervalo  $[0, 1]$ , la integral de una función escalonada con un número finito de discontinuidades,

$$\int_0^1 f(x) dx = \int_{[0,1]} f(x) d\mu(x),$$

corresponde al área bajo la gráfica de  $f$ .

Como ilustración, consideremos la **función de distribución acumulada**, que se define como la probabilidad de que una variable aleatoria  $\xi$  pertenezca al intervalo  $[0, x]$ ; es decir,

$$F(x) = \text{Prob}\{0 \leq \xi \leq x\} = \int_0^x \varrho(s) ds = \int_{[0,x]} \varrho(s) d\mu(s),$$

donde  $\varrho$  es la densidad de probabilidad. Si la densidad es uniforme, entonces  $F(x) = x$ .

Cuando dicha distribución proviene de una función simple cuyos conjuntos son intervalos de la forma  $I_i = [\frac{i-1}{N}, \frac{i}{N}]$  (cada uno de longitud  $1/N$ ) y los coeficientes  $a_i$  satisfacen  $\frac{1}{N} \sum_{i=1}^N a_i = 1$ , la

función de distribución acumulada adopta la forma

$$F(x) = \frac{1}{N} \sum_{i=1}^{k-1} a_i + a_k \left(x - \frac{k-1}{N}\right),$$

donde  $k$  es el índice tal que  $x \in [\frac{k-1}{N}, \frac{k}{N}]$ . Así,  $F(x)$  está compuesta por  $N$  segmentos rectilíneos que unen los puntos dados por la siguiente tabla:

|        |   |         |                 |                       |         |                          |   |
|--------|---|---------|-----------------|-----------------------|---------|--------------------------|---|
| $x$    | 0 | $1/N$   | $2/N$           | $3/N$                 | $\dots$ | $(N-1)/N$                | 1 |
| $F(x)$ | 0 | $a_1/N$ | $(a_1 + a_2)/N$ | $(a_1 + a_2 + a_3)/N$ | $\dots$ | $\sum_{i=1}^{N-1} a_i/N$ | 1 |

Este resultado será empleado más adelante para construir distribuciones de probabilidad acumuladas, en particular para el conjunto de Cantor.

**Definición 2.3.3.** Dado un espacio métrico compacto  $(\mathbf{X}, d)$  con  $\sigma$ -álgebra de Borel  $\mathcal{B}$  y una medida de Borel  $\mu$ , se llama *partición de  $\mathbf{X}$*  a cualquier colección finita de conjuntos de Borel no vacíos  $\{C_1, C_2, \dots, C_M\} \subset \mathcal{B}$  tales que  $\mathbf{X} = \bigcup_{i=1}^M C_i$  y  $C_i \cap C_j = \emptyset$  para  $i \neq j$ .

**Teorema 2.3.1.** Sea  $(\mathbf{X}, d)$  un espacio métrico compacto, sea  $\mathcal{B}$  su  $\sigma$ -álgebra de Borel,  $\mu$  una medida de Borel y  $f : \mathbf{X} \rightarrow \mathbb{R}$  una función continua. Entonces:

1. Para cada entero positivo  $n$  existe una partición  $\mathcal{Q}_n = \{Q_{n,1}, Q_{n,2}, \dots, Q_{n,M(n)}\}$  de  $\mathbf{X}$  con la propiedad de que el diámetro de cada  $Q_{n,m}$  no excede  $1/n$ .
2. Eligiendo puntos muestrales  $\xi_{n,m} \in Q_{n,m}$ , la sucesión de funciones simples

$$f_n(x) = \sum_{m=1}^{M(n)} f(\xi_{n,m}) \mathbf{1}_{Q_{n,m}}(x), \quad n = 1, 2, 3, \dots$$

converge uniformemente a  $f(x)$ .

3. La sucesión de integrales  $\left\{ \int_{\mathbf{X}} f_n d\mu \right\}_{n=1}^{\infty}$  es convergente.
4. El valor del límite no depende de la elección de las particiones  $\mathcal{Q}_n$  ni de los puntos  $\xi_{n,m}$ .

**Definición 2.3.4.** El límite cuya existencia asegura el Teorema 2.3.1 se denomina *integral de  $f$  con respecto a  $\mu$*  y se denota por

$$\int_{\mathbf{X}} f d\mu = \lim_{n \rightarrow \infty} \int_{\mathbf{X}} f_n d\mu.$$

## 2.4. Densidades Invariantes en Sistemas Dinámicos

En la Sección 2.1 se analizaron las frecuencias relativas generadas por un IFS con condición inicial aleatoria. Ampliemos ahora el estudio a un conjunto (ensemble) de condiciones iniciales  $x_0$  y examinemos la frecuencia con la que un valor inicial cae dentro de un subconjunto  $A \subseteq X$ .

El espacio  $X$  puede interpretarse, por ejemplo, como el espacio de fases de un sistema de partículas, lo que proporciona un modelo físico de sistema dinámico. Una medida de la forma

$$\nu_0(A) = \int_A \varrho_0(x) dx \quad (2.1)$$

se denomina *medida de probabilidad* de  $A$ , donde  $\varrho_0$  es una función no negativa e integrable llamada *densidad de probabilidad*. Observemos que la medida actúa sobre conjuntos, mientras que la densidad es una función de las coordenadas del espacio de fases.

La función  $\varrho_0(x)$  puede ser continua o presentar singularidades; en este último caso, la función de distribución acumulada no admite una densidad ordinaria. Por ello es frecuente utilizar la notación equivalente

$$\nu_0(A) = \int_A d\nu_0(x).$$

Dada una transformación  $f$ , interesa el comportamiento de las órbitas para distintas condiciones iniciales. Sea  $\nu_n$  la distribución de probabilidad en la  $n$ -ésima iteración, de modo que  $\nu_n(A)$  representa la probabilidad de que  $x_n$  pertenezca a  $A$ . Si  $\nu_n$  posee densidad  $\varrho_n$ , entonces se verifica

$$\nu_{n+1}(A) = \nu_n(f^{-1}(A)), \quad (2.2)$$

donde  $f^{-1}(A)$  denota la preimagen de  $A$ . Así, la frecuencia relativa de  $x_{n+1}$  en  $A$  coincide con la de  $x_n$  en  $f^{-1}(A)$ .

Una distribución que no cambia bajo la acción de  $f$  se dice *invariante*; en tal caso,

$$\nu_{n+1}(A) = \nu_n(A). \quad (2.3)$$

Una medida que satisface simultáneamente (2.2) y (2.3) recibe el nombre de *medida invariante*. La densidad asociada, que denotaremos por  $\varrho$ , describe un conjunto de puntos cuya distribución no varía con el tiempo; en el espacio de fases representa el estado estacionario, mientras que en el contexto de fractales corresponde al atractor del sistema.

Para una medida invariante  $\nu$  con densidad  $\varrho$ , la relación (2.2) se traduce en

$$\int_A \varrho(x) dx = \int_{f^{-1}(A)} \varrho(x) dx.$$

El *valor esperado* de un observable  $Q(x)$  respecto a la densidad invariante  $\varrho$  se define como

$$\langle Q \rangle = \int_X \varrho(x) Q(x) dx = \int_X Q(x) d\nu(x).$$

De la invarianza se deduce que  $\langle Q \rangle = \langle Q \circ f \rangle$ .

El *promedio temporal* de  $Q$  a lo largo de una órbita se define mediante

$$\overline{Q} = \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{n=0}^{N-1} Q(x_n).$$

En el Capítulo 1 se introdujo el concepto de transformación *mezclante*, que bajo iteraciones sucesivas distribuye uniformemente los conjuntos en el espacio. Tales transformaciones poseen órbitas que visitan toda la región accesible, denominándose *órbitas ergódicas*. Para una densidad inicial suave  $\varrho_0$ , si  $f$  es mezclante se cumple

$$\lim_{n \rightarrow \infty} \varrho_n(x) = \varrho(x),$$

donde  $\varrho$  es la *densidad invariante natural*, que puede representarse formalmente como

$$\lim_{N \rightarrow \infty} \frac{1}{N} \sum_{n=0}^{N-1} \delta(x - x_n).$$

Toda transformación mezclante es ergódica, pero la recíproca no es cierta.

**Definición 2.4.1** (Ergodicidad). *Una transformación junto con una medida invariante se dice ergódica si para toda función de prueba  $Q(x)$  el promedio temporal coincide con el valor esperado:*

$$\overline{Q} = \langle Q \rangle.$$

La ergodicidad implica que el promedio temporal es independiente de la condición inicial.

### 2.4.1. Conjugación de Densidades

En el Capítulo 1 se definió la conjugación topológica entre espacios métricos. Apliquemos ahora esta idea a las densidades de probabilidad.

El mapa de Ulam (logístico con parámetro  $\lambda = 2$ ) es topológicamente conjugado con el mapa de la tienda. Mediante el cambio de variable

$$\theta(y) = -\cos(\pi y) = x,$$

se obtiene la transformación conjugada  $g(y)$ , que resulta ser el mapa de la tienda. La densidad invariante  $\varrho_x(x)$  de  $f(x)$  se relaciona con la densidad  $\varrho_y(y)$  de  $g(y)$  mediante

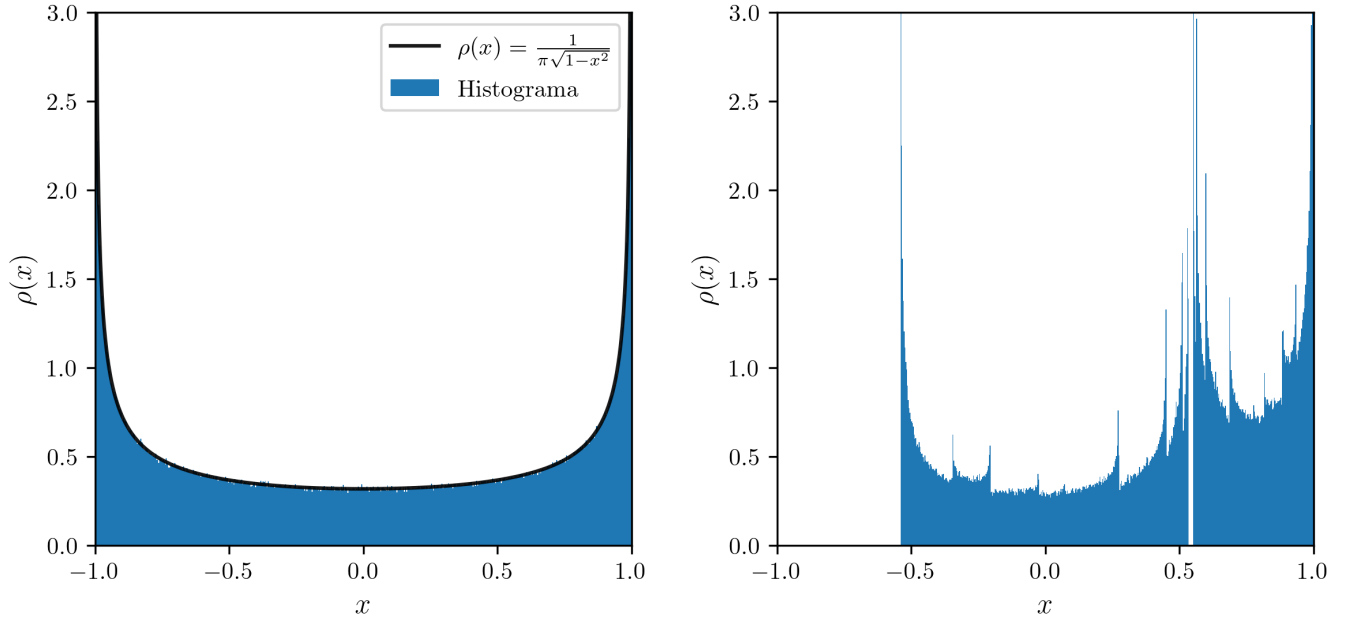
$$\varrho_y(y) = \varrho_x(x) |\det D\theta(y)|, \tag{2.4}$$

expresión que refleja la conservación de la probabilidad:  $\varrho_y(y) dy = \varrho_x(x) dx$ .

Para el mapa de la tienda,  $\varrho_y(y) = 1$ . Derivando la relación  $y = \frac{1}{\pi} \arccos(-x)$  se obtiene la densidad invariante del mapa de Ulam:

$$\varrho(x) = \frac{1}{\pi \sqrt{1-x^2}}.$$

La Figura 2.3 compara histogramas normalizados del mapa de Ulam y del mapa logístico con  $\lambda = 1.54$  con sus respectivas densidades invariantes.



**Figura 2.3:** Izquierda: histograma normalizado del mapa de Ulam (un millón de puntos, ancho de bin 0.001) superpuesto a su densidad invariante. Derecha: histograma análogo para el mapa logístico con  $\lambda = 1.54$ .

## 2.5. El Operador de Perron-Frobenius

La evolución temporal de las densidades de probabilidad inducida por una transformación  $f$  puede describirse mediante un operador lineal.

**Definición 2.5.1** (Operador de Perron-Frobenius). *Sea  $\varrho_n$  la densidad asociada a la medida  $\nu_n$ . El operador lineal  $\mathcal{P}$  que satisface*

$$\mathcal{P}\varrho_n = \varrho_{n+1}$$

*se denomina operador de Perron-Frobenius. Dicho operador gobierna la evolución de la densidad.*

En el caso unidimensional, para un intervalo  $A = [a, b]$ , partiendo de (2.2) se obtiene

$$\int_a^b \varrho_{n+1}(y) dy = \sum_{\sigma} \int_{I_{\sigma}} \varrho_n(x) \mathbf{1}_{[a,b]}(f(x)) dx,$$

donde  $I_{\sigma}$  son los subintervalos en los que  $f$  es monótona e invertible, y  $\mathbf{1}_{[a,b]}$  denota la función característica del intervalo  $[a, b]$ .

Restringiendo a cada rama e introduciendo las inversas locales  $f_{\sigma}^{-1}$ , se obtiene

$$\int_a^b \varrho_{n+1}(y) dy = \sum_{\sigma} \int_{f_{\sigma}^{-1}(a)}^{f_{\sigma}^{-1}(b)} \varrho_n(x) dx,$$

siempre que  $[a, b] \subset f_{\sigma}(I_{\sigma})$ .

Efectuando el cambio de variable  $x = f_{\sigma}^{-1}(y)$ , llegamos a

$$\varrho_{n+1}(y) = \sum_{\sigma} \frac{\varrho_n(f_{\sigma}^{-1}(y))}{|f'_{\sigma}(f_{\sigma}^{-1}(y))|} \mathbf{1}_{f(I_{\sigma})}(y).$$

Por consiguiente, la acción del operador puede escribirse como

$$\mathcal{P}\varrho(y) = \sum_{\sigma} \frac{\varrho(f_{\sigma}^{-1}(y))}{|f'_{\sigma}(f_{\sigma}^{-1}(y))|} \mathbf{1}_{f(I_{\sigma})}(y) = \sum_{x \in f^{-1}(y)} \frac{\varrho(x)}{|f'(x)|}. \quad (2.5)$$

Una representación integral equivalente es

$$\mathcal{P}\varrho(y) = \int_X \varrho(x) \delta(y - f(x)) dx.$$

En dimensión  $d$ , la fórmula general se convierte en

$$\mathcal{P}\varrho(y) = \sum_{x \in f^{-1}(y)} \frac{\varrho(x)}{|\det J_f(x)|},$$

siendo  $J_f(x)$  la matriz jacobiana de  $f$  en el punto  $x$ .

### 2.5.1. Puntos Fijos del Operador

Una función  $\varphi$  que satisface  $\mathcal{P}\varphi = \varphi$  se denomina *función invariante* (punto fijo del operador). Las densidades invariantes son casos particulares de funciones invariantes.

#### Ejemplo: Mapa de la Tienda

Para  $g(x)$  definido por

$$g(x) = \begin{cases} 2x, & 0 \leq x \leq \frac{1}{2}, \\ 2(1-x), & \frac{1}{2} < x \leq 1, \end{cases}$$

cada  $y$  posee dos preimágenes:  $x_1 = y/2$  y  $x_2 = 1 - y/2$ , con  $|g'(x)| = 2$ . La ecuación  $\mathcal{P}\varrho = \varrho$  se convierte en

$$\varrho(y) = \frac{1}{2}\varrho\left(\frac{y}{2}\right) + \frac{1}{2}\varrho\left(1 - \frac{y}{2}\right).$$

La solución normalizada es  $\varrho(y) = 1$ .

#### Ejemplo: Mapa de Ulam

Para  $f(x) = 1 - 2x^2$  en  $[-1, 1]$ , se tiene  $|f'(x)| = 4|x|$  y las preimágenes son  $f_{\pm}^{-1}(y) = \pm\sqrt{(1-y)/2}$ . La ecuación  $\mathcal{P}\varrho = \varrho$  adopta la forma

$$\varrho(y) = \frac{1}{4\sqrt{2(1-y)}} \left[ \varrho(\sqrt{(1-y)/2}) + \varrho(-\sqrt{(1-y)/2}) \right].$$

Puede verificarse que  $\varrho(y) = 1/(\pi\sqrt{1-y^2})$  satisface esta ecuación.

En general, encontrar puntos fijos del operador de Perron-Frobenius es una tarea difícil, y solo se conocen soluciones analíticas para un reducido número de aplicaciones.



### 2.5.2. Espectro del Operador

Las densidades invariantes son funciones propias de  $\mathcal{P}$  asociadas al valor propio 1. En general, las funciones propias satisfacen

$$\mathcal{P}\varphi_\alpha = \lambda_\alpha \varphi_\alpha,$$

donde  $\varphi_\alpha$  pertenece a un espacio funcional adecuado y  $\lambda_\alpha$  es el valor propio correspondiente.

La densidad invariante natural  $\varphi_0$  corresponde al valor propio dominante  $\lambda_0 = 1$ . El segundo valor propio en módulo,  $\lambda_1$ , determina la velocidad de aproximación al estado estacionario. Suponiendo que existe un conjunto completo de funciones propias y que hay un hueco espectral entre  $\lambda_0$  y  $\lambda_1$ , una densidad inicial  $\varrho_0 = \sum_\alpha c_\alpha \varphi_\alpha$  evoluciona como

$$\varrho_N = \mathcal{P}^N \varrho_0 = \sum_\alpha \lambda_\alpha^N c_\alpha \varphi_\alpha = c_0 \varphi_0 + R_N,$$

con  $R_N \rightarrow 0$  cuando  $N \rightarrow \infty$ , dado que  $|\lambda_1| < 1$ . El residuo decae asintóticamente como  $R_N \sim |\lambda_1|^N$ .

Si la transformación es **exponencialmente mezclante**, la convergencia es exponencial:

$$\|\varrho_N - \varphi_0\| \sim e^{-\gamma N} = |\lambda_1|^N,$$

donde  $\gamma^{-1} > 0$  se denomina *tiempo de relajación*.

## 2.6. Operador de Koopman

Mientras que el operador de Perron-Frobenius describe la evolución de las densidades de probabilidad bajo la acción de una transformación  $f$ , existe un operador dual que actúa sobre las funciones observables del sistema. Este operador, introducido por Bernard Koopman en 1931 [28], proporciona una perspectiva complementaria para el estudio de sistemas dinámicos que ha ganado renovada importancia en aplicaciones computacionales modernas [29, 30].

**Definición 2.6.1** (Operador de Koopman). *Sea  $(\mathbf{X}, d)$  un espacio métrico y  $f : \mathbf{X} \rightarrow \mathbf{X}$  una transformación medible. El operador de Koopman  $U$  actúa sobre el espacio de funciones observables  $g : \mathbf{X} \rightarrow \mathbb{C}$  mediante:*

$$Ug(x) = g(f(x)). \tag{2.6}$$

*Es decir,  $U$  avanza el observable  $g$  en un paso temporal mediante la composición con la transformación  $f$ .*

A diferencia del operador de Perron-Frobenius que actúa sobre densidades (elementos del espacio dual), el operador de Koopman actúa directamente sobre las funciones observables. Esta dualidad se manifiesta en la relación fundamental [31]:

$$\langle L\rho, g \rangle = \langle \rho, Ug \rangle,$$

donde  $\langle \cdot, \cdot \rangle$  denota el producto interno en  $L^2(\mathbf{X}, \mu)$  para una medida invariante  $\mu$ .

**Definición 2.6.2** (Eigenfunción de Koopman en Tiempo Discreto). *Una eigenfunción de Koopman  $\phi(\mathbf{x})$  en tiempo discreto, correspondiente al eigenvalor  $\lambda$ , satisface:*

$$\phi(\mathbf{x}_{k+1}) = U_{\Delta t} \phi(\mathbf{x}_k) = \lambda \phi(\mathbf{x}_k) \quad (2.7)$$

donde  $U_{\Delta t}$  es el operador de Koopman para un paso temporal  $\Delta t$ .

**Definición 2.6.3** (Eigenfunción de Koopman en Tiempo Continuo). *En tiempo continuo, una eigenfunción de Koopman  $\phi(\mathbf{x})$  satisface:*

$$\frac{d}{dt} \phi(\mathbf{x}) = U \phi(\mathbf{x}) = \lambda \phi(\mathbf{x}) \quad (2.8)$$

La identificación de eigenfunciones de Koopman a partir de datos o expresiones analíticas constituye un desafío central en los sistemas dinámicos modernos. El descubrimiento de estas eigenfunciones permite representaciones lineales globales de sistemas fuertemente no lineales [32].

### 2.6.1. Propiedades del Operador de Koopman

El operador de Koopman posee varias propiedades fundamentales que lo hacen particularmente útil en aplicaciones [32]:

1. **Linealidad:** Para cualesquiera funciones  $g, h$  y escalares  $\alpha, \beta \in \mathbb{C}$ ,

$$U(\alpha g + \beta h) = \alpha U g + \beta U h.$$

Esta linealidad contrasta con la naturaleza frecuentemente no lineal de la transformación subyacente  $f$ .

2. **Preservación de constantes:** Para cualquier función constante  $g(x) = c$ ,

$$U g(x) = c.$$

3. **Propiedad de composición:** Para iteraciones sucesivas,

$$U^n g(x) = g(f^{\circ n}(x)).$$

La linealidad del operador de Koopman constituye una ventaja significativa, pues permite aplicar las herramientas del análisis funcional lineal a sistemas dinámicos no lineales [33].

### 2.6.2. Espectro del Operador de Koopman

Al igual que el operador de Perron-Frobenius, el operador de Koopman posee un espectro que caracteriza propiedades dinámicas importantes. Las funciones propias de  $U$  satisfacen:

$$U\phi(x) = \lambda\phi(x),$$

donde  $\lambda$  es el valor propio y  $\phi$  la función propia correspondiente.

**Teorema 2.6.1** (Dualidad Koopman–Perron-Frobenius). *Sea  $f : X \rightarrow X$  una transformación medible. El operador de Koopman  $U$ , definido por*

$$(Ug)(x) = g(f(x)),$$

*es el operador adjunto del operador de Perron-Frobenius  $\mathcal{P}$  en el sentido de que*

$$\int_X (Ug)(x) \varrho(x) dx = \int_X g(x) (\mathcal{P}\varrho)(x) dx,$$

*para toda función observable  $g$  y toda densidad  $\varrho$  suficientemente regulares.*

*En consecuencia, los operadores  $U$  y  $\mathcal{P}$  son duales, y sus espectros están relacionados a través de esta dualidad funcional.*

Esta relación implica que el operador de Koopman actúa sobre observables, mientras que el operador de Perron-Frobenius actúa sobre densidades de probabilidad. La dualidad entre ambos operadores refleja dos perspectivas complementarias de la dinámica: la evolución de funciones frente a la evolución de medidas.

En general, aunque  $U$  y  $\mathcal{P}$  son operadores adjuntos, sus espectros pueden diferir dependiendo del espacio funcional considerado. Sin embargo, esta dualidad permite transferir información espectral entre ambos operadores en contextos adecuados.

### 2.6.3. Aplicación

Para ilustrar el funcionamiento del operador de Koopman en sistemas dinámicos concretos, presentamos una serie de ejemplos analíticos donde es posible calcular explícitamente las eigenfunciones mediante desarrollos en serie.

#### Dinámica Lineal

Consideremos el sistema dinámico lineal más simple:

$$\frac{dx}{dt} = x \tag{2.9}$$

Suponemos un desarrollo en serie de Taylor para la eigenfunción  $\phi(x)$ :

$$\phi(x) = c_0 + c_1x + c_2x^2 + c_3x^3 + \dots \quad (2.10)$$

El gradiente y la derivada direccional están dados por:

$$\nabla\phi = c_1 + 2c_2x + 3c_3x^2 + 4c_4x^3 + \dots \quad (2.11)$$

$$\nabla\phi \cdot f = c_1x + 2c_2x^2 + 3c_3x^3 + 4c_4x^4 + \dots \quad (2.12)$$

Resolviendo la ecuación de eigenfunciones de Koopman  $\nabla\phi \cdot f = \lambda\phi$ , obtenemos que  $c_0 = 0$ . Para cualquier entero positivo  $\lambda = k \in \mathbb{Z}^+$ , solo un coeficiente puede ser no nulo. Específicamente, para  $\lambda = k$ , la eigenfunción es  $\phi(x) = cx^k$  para cualquier constante  $c$ .

### Dinámica No Lineal Cuadrática

Consideremos ahora un sistema no lineal:

$$\frac{dx}{dt} = x^2 \quad (2.13)$$

En este caso, no existe desarrollo en serie de Taylor que satisfaga la ecuación de eigenfunciones, excepto la solución trivial  $\phi = 0$  para  $\lambda = 0$ . En su lugar, asumimos un desarrollo en serie de Laurent:

$$\phi(x) = \dots + c_{-3}x^{-3} + c_{-2}x^{-2} + c_{-1}x^{-1} + c_0 + c_1x + c_2x^2 + c_3x^3 + \dots \quad (2.14)$$

Las derivadas correspondientes son:

$$\nabla\phi = \dots - 3c_{-3}x^{-4} - 2c_{-2}x^{-3} - c_{-1}x^{-2} + c_1 + 2c_2x + 3c_3x^2 + \dots \quad (2.15)$$

$$\nabla\phi \cdot f = \dots - 3c_{-3}x^{-2} - 2c_{-2}x^{-1} - c_{-1} + c_1x^2 + 2c_2x^3 + 3c_3x^4 + \dots \quad (2.16)$$

Resolviendo los coeficientes de la serie de Laurent que satisfacen  $\nabla\phi \cdot f = \lambda\phi$ , encontramos que todos los coeficientes con índice positivo son cero ( $c_k = 0$  para  $k \geq 1$ ). Los coeficientes con índice negativo satisfacen la recursión  $\lambda c_{k+1} = kc_k$  para  $k \leq -1$ . La solución es:

$$\phi(x) = c_0 \left( 1 - \lambda x^{-1} + \frac{\lambda^2}{2}x^{-2} - \frac{\lambda^3}{6}x^{-3} + \dots \right) = c_0 e^{-\lambda/x} \quad (2.17)$$

Esta solución es válida para todos  $\lambda \in \mathbb{C}$ . Para  $\phi(x) = e^{-\lambda/x}$ , tenemos:

$$\nabla\phi = \frac{\lambda}{x^2} e^{-\lambda/x} \quad (2.18)$$

$$\nabla\phi \cdot f = \frac{\lambda}{x^2} e^{-\lambda/x} \cdot x^2 = \lambda e^{-\lambda/x} = \lambda\phi(x) \quad (2.19)$$

## Dinámica No Lineal Polinomial General

Para un sistema dinámico más general:

$$\frac{dx}{dt} = ax^n \quad (2.20)$$

se puede verificar que la función:

$$\phi(x) = e^{\frac{\lambda x^{1-n}}{(1-n)a}} \quad (2.21)$$

es una eigenfunción para todo  $\lambda \in \mathbb{C}$ .

$$\nabla \phi = \frac{\lambda}{a} x^{-n} e^{\frac{\lambda x^{1-n}}{(1-n)a}} \quad (2.22)$$

$$\nabla \phi \cdot f = \frac{\lambda}{a} x^{-n} e^{\frac{\lambda x^{1-n}}{(1-n)a}} \cdot ax^n = \lambda e^{\frac{\lambda x^{1-n}}{(1-n)a}} = \lambda \phi(x) \quad (2.23)$$

## Mapa de la Tienda

Para el mapa de la tienda  $g(x)$ , consideremos observables en  $L^2([0, 1], dx)$ . Las funciones propias del operador de Koopman satisfacen:

$$U\phi(x) = \phi(g(x)) = \lambda\phi(x).$$

Para la función constante  $\phi_0(x) = 1$ , se tiene:

$$U\phi_0(x) = \phi_0(g(x)) = 1 = \lambda\phi_0(x) \Rightarrow \lambda = 1.$$

Pueden encontrarse otras funciones propias, como  $\phi_1(x) = \cos(\pi x)$ , para la cual:

$$U\phi_1(x) = \cos(\pi g(x)) = \cos(2\pi x) = \lambda \cos(\pi x).$$

Sin embargo,  $\cos(2\pi x)$  no es múltiplo de  $\cos(\pi x)$ .

Para el mapa de Ulam  $f(x) = 1 - 2x^2$  en  $[-1, 1]$ , el operador de Koopman actúa como:

$$Ug(x) = g(1 - 2x^2).$$

Utilizando la conjugación con el mapa de la tienda mediante  $\theta(y) = -\cos(\pi y)$ , podemos relacionar los operadores de Koopman de ambos sistemas. Si  $\psi$  es una función propia de  $U$  para el mapa de Ulam, entonces:

$$\tilde{\psi}(y) = \psi(\theta(y))$$

es una función propia del operador de Koopman para el mapa de la tienda.

En ambos mapas, el de la tienda y el de Ulam, no existen autofunciones no constantes del operador de Koopman  $U$ . Para ilustrarlo, consideremos el mapa de duplicación  $g(x) = 2x \bmod 1$  y la función base exponencial  $f(x) = e^{2\pi i x}$ . Bajo la iteración de  $U$ , se obtiene:

$$U^n f(x) = e^{2\pi i 2^n x},$$

y las funciones  $\{U^n f\}_{n \geq 0}$  resultan ortogonales entre sí en  $L^2([0, 1])$ . En el subespacio generado por estas funciones, el operador  $U$  actúa como un *operador de desplazamiento a la derecha* sobre una secuencia infinita, el cual carece de autovalores distintos de cero.

De hecho, como demuestran Otto y Rowley [34], en el espacio completo  $L^2([0, 1])$  el único autovalor de  $U$  es  $\lambda = 1$ , correspondiente a la función constante. El resto del espectro de  $U$  no es puntual, sino *continuo*, distribuyéndose sobre el círculo unitario en el plano complejo.

Más generalmente, un resultado clásico de la teoría ergódica establece que un sistema dinámico es *débilmente mezclante* (*weakly mixing*) si y sólo si su operador de Koopman no posee autofunciones no constantes. Así, la naturaleza caótica y mezclante de los mapas de la tienda y de Ulam se refleja directamente en la estructura espectral de  $U$ , caracterizada por un espectro puramente continuo y la ausencia de subespacios invariantes de dimensión finita que simplifiquen la dinámica.

Por último, cabe destacar que esta carencia de autofunciones no triviales implica que no existen coordenadas globales capaces de linealizar la dinámica del sistema. En otras palabras, no hay un cambio de variable que transforme la evolución no lineal en un sistema lineal equivalente. No obstante, esto no invalida los métodos basados en el operador de Koopman: éstos pueden utilizarse con éxito para la predicción a corto plazo y el análisis modal de sistemas caóticos, incluso en ausencia de un espectro puntual.

#### 2.6.4. Descomposición Modal Dinámica

Una aplicación poderosa del operador de Koopman es la *descomposición modal dinámica* (DMD), desarrollada inicialmente en el contexto de dinámica de fluidos [35] y posteriormente generalizada [36]:

**Teorema 2.6.2** (Descomposición Modal). *Bajo condiciones adecuadas, un observable  $g$  puede expresarse como:*

$$g(x) = \sum_{\alpha} c_{\alpha} \phi_{\alpha}(x),$$

donde  $\phi_{\alpha}$  son funciones propias de  $U$ . Entonces, la evolución temporal está dada por:

$$U^n g(x) = \sum_{\alpha} c_{\alpha} \lambda_{\alpha}^n \phi_{\alpha}(x).$$

Esta descomposición es particularmente útil para identificar comportamientos asintóticos, donde los valores propios con mayor magnitud dominan la dinámica a largo plazo [37].

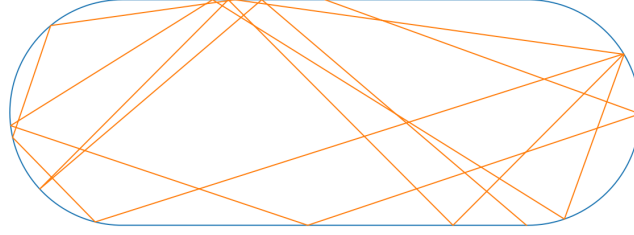
#### 2.6.5. Relación con la Ergodicidad

El operador de Koopman proporciona una caracterización elegante de la ergodicidad [33]:

**Teorema 2.6.3** (Ergodicidad y Espectro). *Una transformación  $f$  con medida invariante  $\mu$  es ergódica si y solo si  $\lambda = 1$  es un valor propio simple del operador de Koopman, es decir, si las únicas funciones propias con valor propio 1 son las funciones constantes.*

Esta caracterización conecta directamente con la definición de ergodicidad presentada anteriormente, donde el promedio temporal coincide con el promedio espacial para observables arbitrarios.

### 2.6.6. Ejemplo de Aplicación en Billares Clásicos



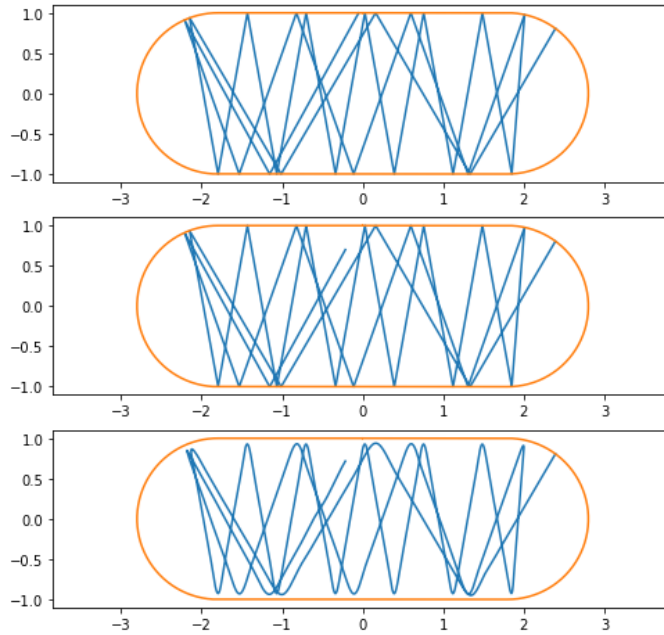
**Figura 2.4:** Trayectoria de una partícula moviéndose dentro de la curva cerrada con forma de estadio.

El movimiento de una partícula dentro de un billar clásico está compuesto por trayectorias rectilíneas que se interrumpen al chocar con las paredes del billar. En el espacio de configuraciones, estas trayectorias se representan como segmentos de recta que conectan los sucesivos puntos de colisión, como se ilustra en la figura 2.5. Por consiguiente, el problema se reduce a determinar los puntos de impacto sobre la frontera y los ángulos de reflexión, lo cual equivale a encontrar la dirección del vector velocidad de la partícula. En el espacio de fase, se define una sección de Poincaré [38, 39] mediante la longitud de arco a lo largo de la frontera y la proyección de la velocidad sobre el vector tangente a la frontera inmediatamente después de cada colisión [40]. El comportamiento del sistema depende completamente de la geometría del billar, pudiendo ser integrable (como en los casos rectangular o elíptico) o caótico (como en los billares de Sinai, cardioides o estadios).

### Aplicación de DMD a Billares Clásicos

Las trayectorias de una partícula libre dentro de un billar pueden aproximarse localmente en espacio y tiempo mediante la descomposición en modos dinámicos (DMD). Como ejemplo, consideremos un billar con forma de estadio como el mostrado en la figura 2.5. Para este estudio, se simuló cinco mil trayectorias partiendo de un mismo punto seleccionado aleatoriamente, con direcciones iniciales también aleatorias y magnitud de velocidad unitaria. Se construyó una matriz  $\mathbf{X}$  compuesta por los vectores  $\mathbf{x}_k$ , que contienen las posiciones de las trayectorias en tiempos discretos  $t_k = k\Delta t$ , con  $\Delta t = 0.1$  hasta un tiempo final  $t_m = 50$  (correspondiente a  $m = 500$  instantes).

La figura 2.5 muestra los resultados para una de las trayectorias analizadas con DMD. En la parte superior se presenta la trayectoria construida mediante la conexión de los puntos de colisión sobre la frontera. En la parte media se muestra la trayectoria formada por las posiciones sucesivas de la partícula en intervalos  $\Delta t$ . Finalmente, en la parte inferior se exhibe la trayectoria reconstruida a partir de los modos dinámicos.



**Figura 2.5:** Ejemplo de aplicación de DMD a un billar con forma de estadio.

Si bien este ejemplo demuestra la viabilidad de descomponer las trayectorias en modos dinámicos, para obtener aproximaciones más precisas es necesario considerar valores mayores de  $r$  (número de modos retenidos). No obstante, esto puede complicar el análisis de los patrones presentes en los modos dinámicos. A pesar de esta limitación, consideramos que el método DMD, basado en datos experimentales o de simulación, constituye una herramienta computacional prometedora para el análisis de billares clásicos.

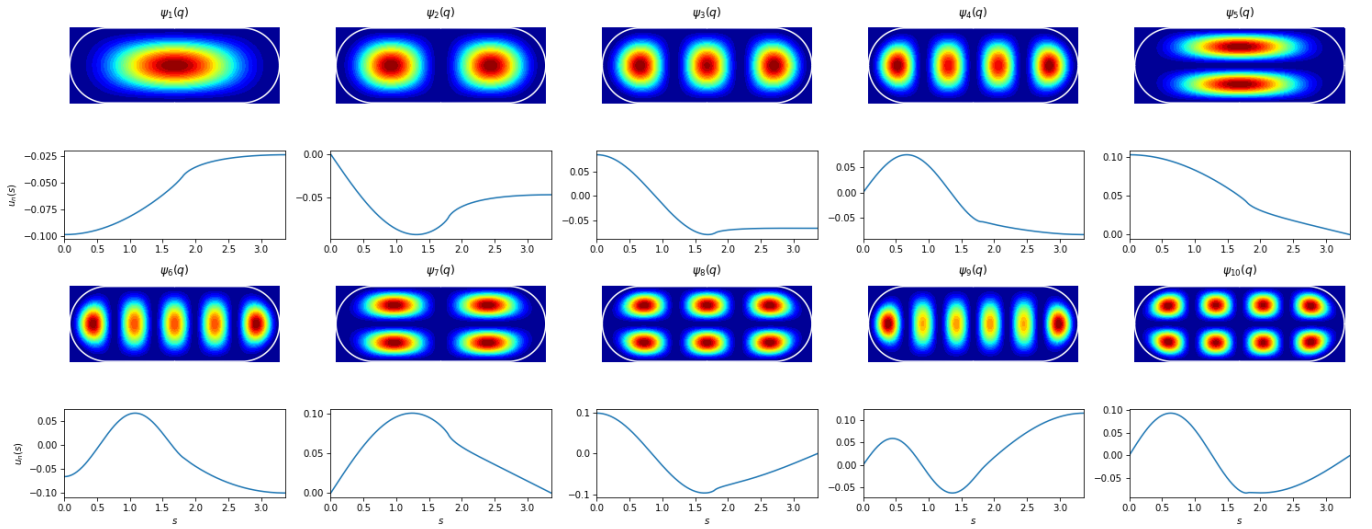
### 2.6.7. Comparación con Métodos Numéricos para Billares Cuánticos

La aproximación numérica para la determinación de valores propios en billares cuánticos, como la desarrollada por Bäcker (2002) [41], presenta similitudes fundamentales con el enfoque DMD en cuanto a su estructura matemática subyacente. Ambos métodos recurren a técnicas de álgebra lineal numérica para resolver problemas complejos: mientras el método de Bäcker utiliza la descomposición en valores singulares (SVD) para encontrar los ceros del determinante de la matriz  $A_k$  en la ecuación integral de frontera, DMD emplea SVD para descomponer la matriz de datos y aproximar el operador de Koopman. Esta convergencia metodológica refleja cómo problemas aparentemente distintos -la cuantización de sistemas caóticos y la descomposición modal de dinámicas no lineales- comparten



una base matemática común en el análisis espectral de operadores lineales.

La figura 2.6 ilustra esta conexión desde una perspectiva física, mostrando las diez primeras funciones de onda obtenidas mediante métodos numéricos similares a los descritos por Bäcker. Estos estados estacionarios, caracterizados por sus patrones nodales específicos, pueden verse como análogos cuánticos de los modos dinámicos identificados por DMD en sistemas clásicos. La principal diferencia reside en que los modos DMD capturan evolución temporal en sistemas clásicos, mientras que las autofunciones cuánticas representan estados estacionarios con energías bien definidas. No obstante, ambos enfoques comparten el objetivo de descomponer sistemas complejos en componentes fundamentales que revelen su estructura subyacente, demostrando la utilidad de técnicas numéricas avanzadas en la caracterización de sistemas dinámicos tanto clásicos como cuánticos.



**Figura 2.6:** Funciones de onda  $|\psi_n(q)|^2$  para los diez primeros estados de energía. En la parte de abajo se muestran las respectivas derivadas normales  $u_n(s)$  en el primer cuarto del estadio.

### 2.6.8. Ventajas Computacionales

El operador de Koopman ofrece ventajas computacionales significativas que han impulsado su uso en aplicaciones modernas [30]:

- **Linealidad:** Aunque el sistema dinámico subyacente sea no lineal, el operador de Koopman es lineal, permitiendo el uso de técnicas de álgebra lineal numérica.
- **Análisis de datos:** La DMD puede aplicarse directamente a datos experimentales para identificar modos dinámicos dominantes sin conocimiento previo de las ecuaciones del sistema [35].
- **Predicción:** Una vez identificados los modos de Koopman, la predicción de la evolución temporal se reduce a la evolución de una combinación lineal de funciones propias [37].

## 2.7. El Operador de Markov como Transformación Contractiva en el Espacio de Medidas

Sea  $(\mathbf{X}, d)$  un espacio métrico completo y sea  $\mathcal{B}$  su  $\sigma$ -álgebra de Borel. Dada una aplicación continua  $w : \mathbf{X} \rightarrow \mathbf{X}$ , la operación de preimagen  $w^{-1}$  envía conjuntos de Borel en conjuntos de Borel. En consecuencia, si  $\nu$  es una medida de Borel normalizada en  $\mathbf{X}$ , la medida imagen  $\nu \circ w^{-1}$  resulta también una medida de Borel normalizada. Este hecho permite definir una transformación natural en el espacio de todas las medidas de probabilidad de Borel.

**Definición 2.7.1.** *Supongamos que  $(\mathbf{X}, d)$  es un espacio métrico compacto y sea  $\mathcal{P}(\mathbf{X})$  la familia de todas las medidas de Borel normalizadas sobre  $\mathbf{X}$ . Dado un sistema de funciones iteradas (IFS) con pesos probabilísticos*

$$\{\mathbf{X}; w_1, w_2, \dots, w_N; p_1, p_2, \dots, p_N\},$$

*definimos el operador de Markov  $M : \mathcal{P}(\mathbf{X}) \rightarrow \mathcal{P}(\mathbf{X})$  mediante la expresión*

$$M(\nu) = \sum_{i=1}^N p_i \nu \circ w_i^{-1}, \quad \nu \in \mathcal{P}(\mathbf{X}).$$

**Lema 2.7.1.** *Bajo las mismas hipótesis de la Definición 2.7.1, para cualquier función  $f : \mathbf{X} \rightarrow \mathbb{R}$  que sea simple o continua y para toda medida  $\nu \in \mathcal{P}(\mathbf{X})$ , se verifica*

$$\int_{\mathbf{X}} f d(M(\nu)) = \sum_{i=1}^N p_i \int_{\mathbf{X}} (f \circ w_i) d\nu.$$

**Teorema 2.7.1.** *Sea  $(\mathbf{X}, d)$  un espacio métrico compacto y considérese un IFS con probabilidades*

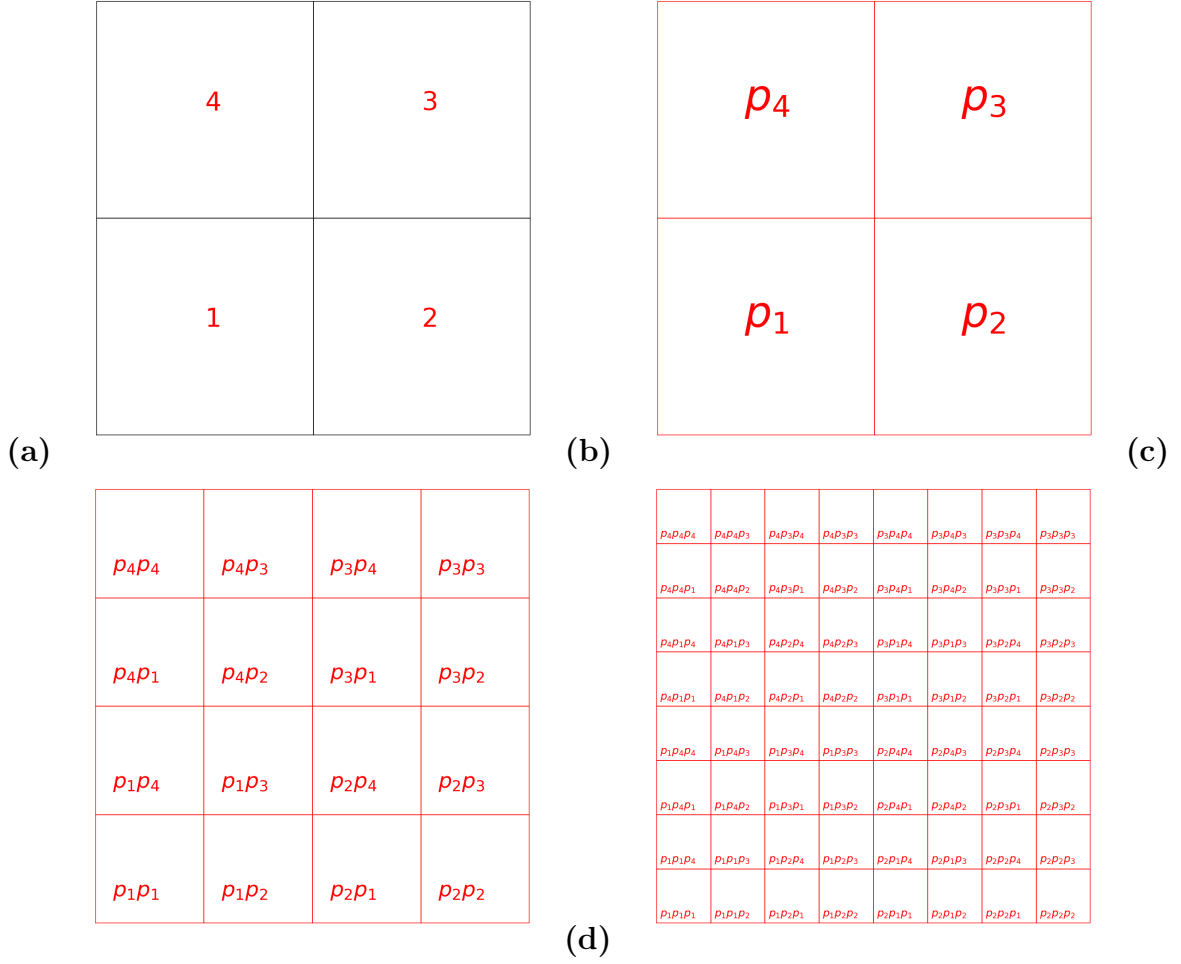
$$\{\mathbf{X}; w_1, w_2, \dots, w_N; p_1, p_2, \dots, p_N\}.$$

*Denotemos por  $s \in (0, 1)$  una constante de contractividad para las transformaciones  $w_i$  y sea  $M$  el operador de Markov correspondiente. Entonces,  $M$  resulta ser una contracción de factor  $s$  en el espacio métrico  $(\mathcal{P}(\mathbf{X}), d_H)$ , donde  $d_H$  es la distancia de Hutchinson. En otras palabras,*

$$d_H(M(\nu), M(\mu)) \leq s d_H(\nu, \mu), \quad \forall \nu, \mu \in \mathcal{P}(\mathbf{X}).$$

*Como consecuencia, existe una única medida  $\mu_* \in \mathcal{P}(\mathbf{X})$  que satisface  $M(\mu_*) = \mu_*$ .*

**Definición 2.7.2.** *La medida  $\mu_*$  que cumple  $\mu_* = \sum_{i=1}^N p_i \mu_* \circ w_i^{-1}$  y que constituye el único punto fijo del operador  $M$  se denomina medida invariante del IFS con probabilidades.*



**Figura 2.7:** (a) Representación de un collage generado mediante un IFS de cuatro aplicaciones. (b), (c) y (d) Distribución de la medida en las distintas regiones tras una, dos y tres aplicaciones sucesivas del operador de Markov.

### 2.7.1. Interpretación Geométrica del Operador de Markov

Para ilustrar el comportamiento del operador de Markov, tomemos un IFS probabilístico definido sobre el cuadrado unitario  $\blacksquare \subset \mathbb{R}^2$ :

$$\{\blacksquare; w_1, w_2, w_3, w_4; p_1, p_2, p_3, p_4\},$$

donde cada  $w_i$  es una transformación contractiva con factor de contracción  $s = 1/4$  y cuyas imágenes cubren  $\blacksquare$ , como se aprecia en el collage de la Figura 2.7(a). Sea  $M$  el operador de Markov correspondiente y elijamos una medida inicial  $\mu_0 \in \mathcal{P}(\blacksquare)$  con  $\mu_0(\blacksquare) = 1$ ; por ejemplo, la medida de Lebesgue normalizada (área).

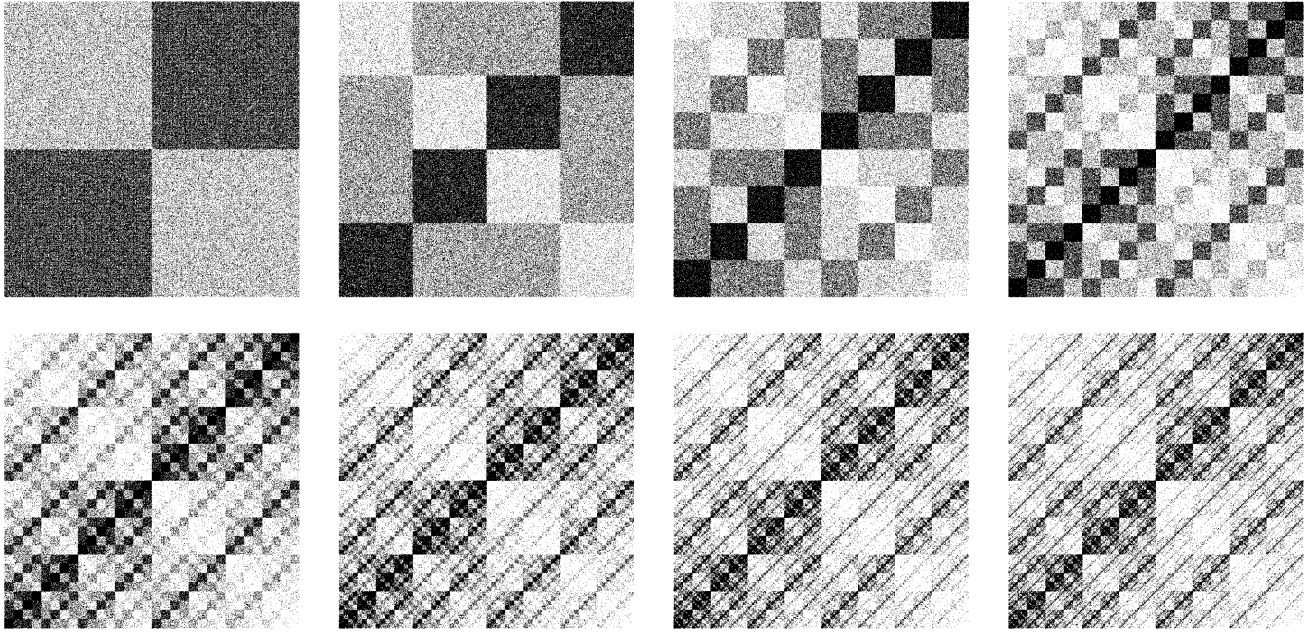
Consideremos la sucesión  $\mu_n = M^n(\mu_0)$  obtenida por iteración repetida de  $M$ . Tras la primera iteración,  $\mu_1 = M(\mu_0)$  asigna a cada región  $w_i(\blacksquare)$  una masa igual a  $p_i$ , es decir,  $\mu_1(w_i(\blacksquare)) = p_i$  para  $i = 1, \dots, 4$  (Figura 2.7(b)). En la segunda iteración,  $\mu_2 = M(\mu_1)$  distribuye la masa de manera que  $\mu_2(w_i \circ w_j(\blacksquare)) = p_i p_j$  para todo  $i, j$  (Figura 2.7(c)). Este patrón continúa: en el paso  $n$ , la medida  $\mu_n$  asigna a cada celda de la forma  $w_{i_1} \circ \dots \circ w_{i_n}(\blacksquare)$  un peso igual al producto  $p_{i_1} \cdots p_{i_n}$ .

En términos intuitivos, cada aplicación de  $M$  redistribuye la “masa” contenida en una celda

## 2.7. EL OPERADOR DE MARKOV COMO TRANSFORMACIÓN CONTRACTIVA EN EL ESPACIO DE MEDIDAS

$\blacksquare_{i_1, \dots, i_k} = w_{i_1} \circ \dots \circ w_{i_k}(\blacksquare)$  entre las cuatro subceldas  $w_1(\blacksquare_{i_1, \dots, i_k})$ ,  $w_2(\blacksquare_{i_1, \dots, i_k})$ ,  $w_3(\blacksquare_{i_1, \dots, i_k})$  y  $w_4(\blacksquare_{i_1, \dots, i_k})$ , respetando las probabilidades  $p_i$ . Simultáneamente, masa proveniente de otras celdas se traslada hacia subceldas de  $\blacksquare_{i_1, \dots, i_k}$ , de modo que la masa total de  $\blacksquare_{i_1, \dots, i_k}$  se preserva. Este mecanismo conduce, en el límite, a la medida invariante  $\mu_*$ .

La Figura 2.8 exhibe ocho instantáneas de dicho proceso: partiendo de una distribución uniforme de puntos (250,000 en total), cada imagen corresponde a una iteración de  $M$  con probabilidades  $(0.4, 0.1, 0.4, 0.1)$ . Nótese cómo la nube de puntos se reacomoda progresivamente, aproximándose al atractor del IFS.



**Figura 2.8:** Ocho etapas sucesivas en la evolución de una medida uniforme bajo la acción del operador de Markov. Cada imagen contiene 250,000 puntos; las probabilidades son  $p_1 = 0.4$ ,  $p_2 = 0.1$ ,  $p_3 = 0.4$ ,  $p_4 = 0.1$ . La distribución se ajusta gradualmente a la medida invariante.

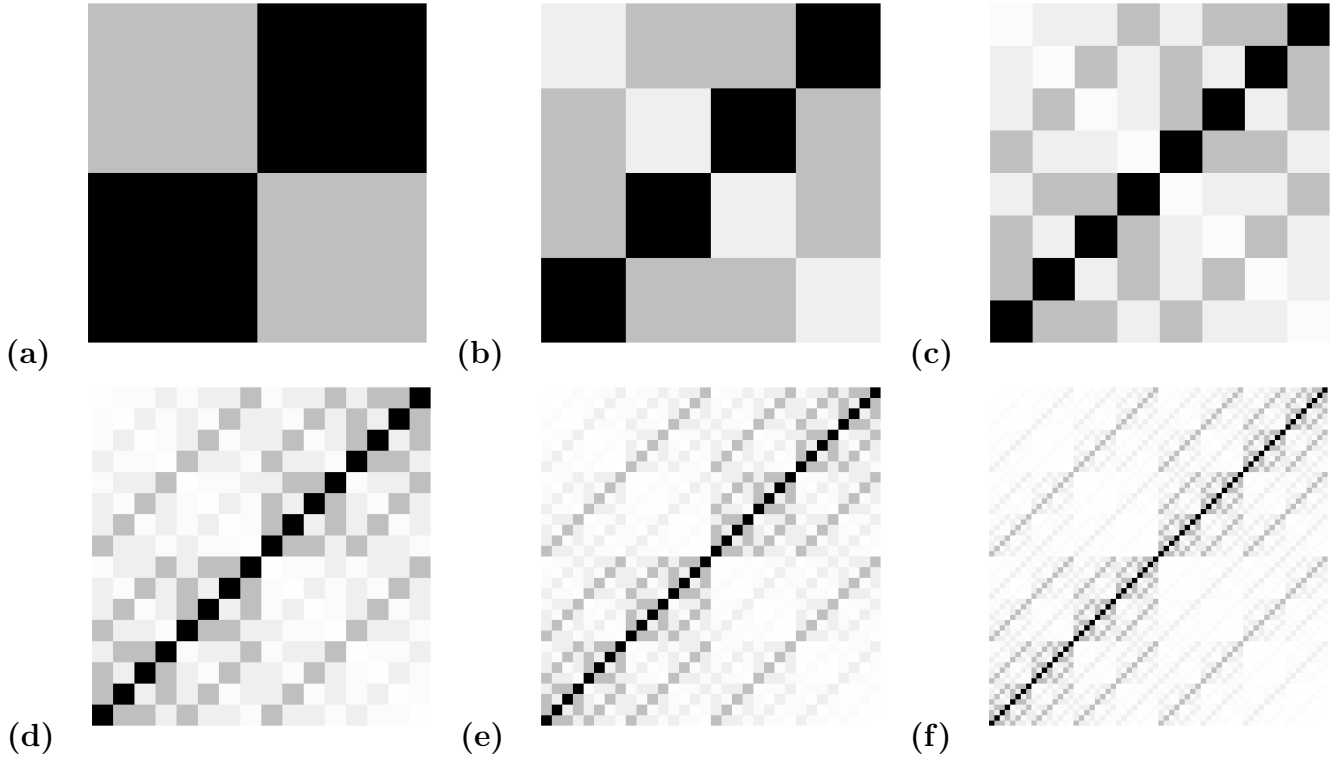
Una representación alternativa, en tonos de gris, se ofrece en la Figura 2.9. Cada panel muestra la densidad de la medida  $\mu_n$  en escala de grises (blanco = masa nula, negro = masa máxima). Las regiones más oscuras indican mayor concentración de probabilidad. La secuencia (a) a (f) corresponde a las primeras seis iteraciones de  $M$ .

### 2.7.2. Propiedades Básicas de la Medida Invariante

**Teorema 2.7.2** (Soporte de la medida invariante). *Sea  $(\mathbf{X}, d)$  un espacio métrico compacto y*

$$\{\mathbf{X}; w_1, w_2, \dots, w_N; p_1, p_2, \dots, p_N\}$$

*un IFS con probabilidades. Si  $\mu_*$  denota la medida invariante de dicho IFS, entonces el soporte de  $\mu_*$  coincide exactamente con el atractor  $A$  del IFS subyacente  $\{\mathbf{X}; w_1, w_2, \dots, w_N\}$ .*



**Figura 2.9:** Evolución de la densidad de la medida bajo iteraciones sucesivas del operador de Markov, representada en escala de grises. Las zonas más oscuras corresponden a mayor probabilidad.

**Teorema 2.7.3** (Teorema del Collage para medidas). *Bajo las mismas hipótesis anteriores, sea  $s \in (0, 1)$  una constante de contracción común para las transformaciones  $w_i$ ,  $M$  el operador de Markov y  $\mu_*$  la medida invariante. Para cualquier medida  $\nu \in \mathcal{P}(\mathbf{X})$  se verifica la estimación*

$$d_H(\nu, \mu_*) \leq \frac{d_H(\nu, M(\nu))}{1 - s}.$$

El Teorema del Collage para medidas proporciona una cota para la distancia entre una medida dada  $\nu$  y la medida invariante  $\mu_*$  en términos de cuánto difiere  $\nu$  de su imagen bajo  $M$ . Este resultado es análogo al teorema del collage para conjuntos visto en el Capítulo 2 y resulta útil en la aproximación práctica de medidas invariantes.

## 2.8. Teorema de Elton y sus Consecuencias

Tras haber sentado los fundamentos de la teoría de medidas y del operador de Markov, estamos en condiciones de presentar un resultado central en el estudio de los IFS con probabilidades: el teorema de Elton. Este teorema establece la conexión entre las órbitas aleatorias generadas por el algoritmo de iteración aleatoria y la medida invariante.

**Teorema 2.8.1** (Teorema de Elton). *Sea  $(\mathbf{X}, d)$  un espacio métrico compacto y consideremos un IFS con probabilidades*

$$\{\mathbf{X}; w_1, w_2, \dots, w_N; p_1, p_2, \dots, p_N\}.$$

Sea  $\{x_n\}_{n=0}^\infty$  una trayectoria obtenida mediante el algoritmo de iteración aleatoria a partir de un punto inicial  $x_0$ ; es decir,

$$x_n = w_{\sigma_n} \circ w_{\sigma_{n-1}} \circ \cdots \circ w_{\sigma_1}(x_0),$$

donde en cada paso el índice  $\sigma_k$  se elige de manera independiente según la distribución  $\{p_i\}$ . Si  $\mu$  denota la única medida invariante del IFS, entonces, con probabilidad uno (respecto a la elección de las sucesiones  $\{\sigma_k\}$ ), se cumple que

$$\lim_{n \rightarrow \infty} \frac{1}{n+1} \sum_{k=0}^n f(x_k) = \int_{\mathbf{X}} f(x) d\mu(x)$$

para cualquier función continua  $f : \mathbf{X} \rightarrow \mathbb{R}$  y para cualquier punto inicial  $x_0$ .

**Corolario 2.8.1.** Sea  $B \in \mathcal{B}(\mathbf{X})$  un conjunto de Borel cuya frontera  $\partial B$  tiene medida nula ( $\mu(\partial B) = 0$ ). Defínase  $\mathcal{N}(B, n)$  como el número de puntos de la sucesión  $\{x_0, x_1, \dots, x_n\}$  que caen en  $B$ . Entonces, con probabilidad uno,

$$\mu(B) = \lim_{n \rightarrow \infty} \frac{\mathcal{N}(B, n)}{n+1}$$

para todo  $x_0$ . En otras palabras, la frecuencia asintótica de visita a  $B$  por la órbita aleatoria coincide con la masa que la medida invariante asigna a  $B$ .

## 2.9. Sistemas de Funciones Iteradas Recurrentes

Hasta ahora hemos examinado dos clases principales de sistemas fractales: aquellos que operan sobre subconjuntos compactos de  $\mathbb{R}^2$  (como el triángulo de Sierpinski) y aquellos que actúan sobre el espacio de medidas de probabilidad  $\mathcal{P}(\mathbb{R}^2)$ . Ambos pueden englobarse dentro del concepto general de *sistema fractal*, el cual consta de tres ingredientes básicos:

1. Un *espacio base*  $(\mathbb{X}, d)$ , que es un espacio métrico (usualmente compacto) donde se sitúan los objetos fractales. Por ejemplo,  $\mathbb{X}$  puede ser  $\blacksquare \subset \mathbb{R}^2$  (el cuadrado unitario).
2. Un *espacio de modelos*  $\mathbb{Y} = \mathbb{Y}(\mathbb{X})$ , cuyos puntos representan los modelos fractales. Este espacio está dotado de una métrica  $h$  que lo hace compacto. Ejemplos típicos son:
  - El espacio  $\mathcal{H}(\mathbb{X})$  de los subconjuntos compactos no vacíos de  $\mathbb{X}$ , con la métrica de Hausdorff  $d_H$ .
  - El espacio  $\mathcal{P}(\mathbb{X})$  de las medidas de Borel normalizadas, con la métrica de Hutchinson  $d_H$ .
  - El espacio  $C[0, 1]$  de las funciones continuas en  $[0, 1]$ , con la métrica del supremo.

3. Un *operador de contracción*  $O : \mathbb{Y} \rightarrow \mathbb{Y}$ , construido a partir de un conjunto finito de aplicaciones contractivas definidas sobre  $\mathbb{X}$ . Este operador satisface

$$h(O(\phi), O(\psi)) \leq s h(\phi, \psi), \quad \forall \phi, \psi \in \mathbb{Y},$$

para alguna constante  $s \in [0, 1)$ .

A continuación se presentan tres ejemplos concretos de sistemas fractales.

**Ejemplo 1 (IFS clásico).**  $\mathbb{X} = \blacksquare \subset \mathbb{R}^2$ ,  $\mathbb{Y} = \mathcal{H}(\mathbb{X})$ . Dadas transformaciones contractivas  $w_i : \mathbb{X} \rightarrow \mathbb{X}$ , el operador de Hutchinson  $W : \mathbb{Y} \rightarrow \mathbb{Y}$  se define por  $W(B) = \bigcup_{i=1}^N w_i(B)$ . Los atractores son los conjuntos fractales usuales.

**Ejemplo 2 (IFS con probabilidades).**  $\mathbb{X} = \blacksquare$ ,  $\mathbb{Y} = \mathcal{P}(\mathbb{X})$ . Con las mismas  $w_i$  y añadiendo probabilidades  $p_i$ , el operador de Markov  $M : \mathbb{Y} \rightarrow \mathbb{Y}$  dado por  $M(\nu) = \sum_i p_i \nu \circ w_i^{-1}$  es una contracción. Sus puntos fijos son medidas invariantes que pueden interpretarse como imágenes en escala de grises.

**Ejemplo 3 (Interpolación fractal).**  $\mathbb{X} = [0, 1]$ ,  $\mathbb{Y} = C[0, 1]$ . Mediante ciertas transformaciones afines se define un operador  $T : \mathbb{Y} \rightarrow \mathbb{Y}$  cuya única función fija es una función fractal de interpolación (ver [2], capítulo 6).

En esta sección introducimos una generalización importante: los *sistemas de funciones iteradas recurrentes* (RIFS, por sus siglas en inglés).

**Definición 2.9.1** (RIFS). *Un sistema de funciones iteradas recurrentes está constituido por:*

- Una IFS  $\{\mathbb{X}; w_1, w_2, \dots, w_N\}$ .
- Una matriz estocástica  $(p_{ij})_{i,j=1}^N$  con entradas en  $[0, 1]$  que cumple:
  1.  $\sum_{j=1}^N p_{ij} = 1$  para cada  $i = 1, \dots, N$ .
  2. Para todo par  $(i, j)$  existe una sucesión finita de índices  $k_1, k_2, \dots, k_m$  tales que

$$p_{ik_1} p_{k_1 k_2} \cdots p_{k_m j} > 0.$$

La matriz  $(p_{ij})$  define una cadena de Markov en tiempo discreto sobre los estados  $\{1, \dots, N\}$ , donde  $p_{ij}$  es la probabilidad de pasar del estado  $i$  al estado  $j$ . La condición (1) garantiza que es una matriz estocástica por filas; la condición (2) asegura que la cadena es irreducible (todo estado es accesible desde cualquier otro).

Nótese que un IFS con probabilidades usual es un caso particular de RIFS: basta tomar  $p_{ij} = p_j$  para todo  $i$ , es decir, la probabilidad de aplicar  $w_j$  no depende del estado anterior.

Un RIFS se dice *hiperbólico* si el IFS subyacente es contractivo (existe  $s \in [0, 1)$  tal que cada  $w_i$  es una contracción de factor  $s$ ). En lo que sigue supondremos siempre hiperbolicidad.

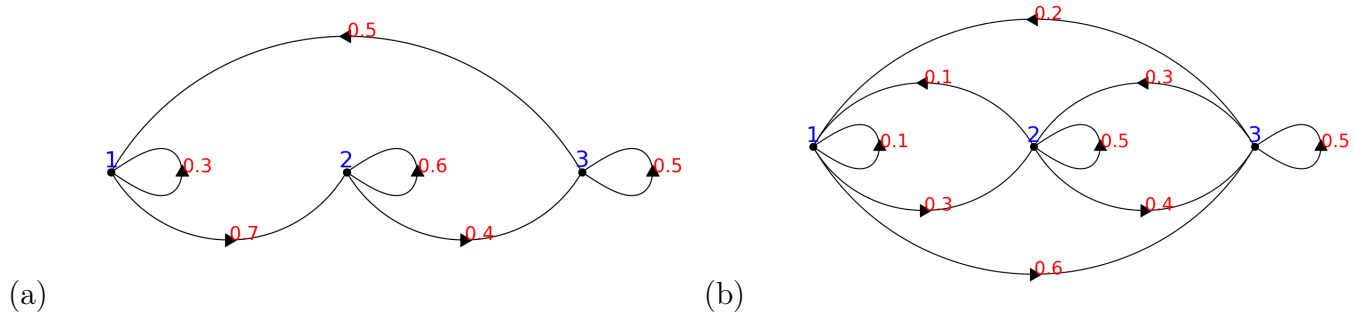
Las Tablas 2.2 y 2.3 muestran dos ejemplos de parámetros para RIFS en  $\mathbb{R}^2$  con transformaciones afines. En el primer caso (Tabla 2.2) algunas transiciones están prohibidas ( $p_{13} = 0$ ,  $p_{21} = 0$ ); en el segundo (Tabla 2.3) todas las transiciones son posibles en un solo paso. Los diagramas de la Figura 2.10 ilustran las estructuras de las cadenas de Markov correspondientes.

**Tabla 2.2:** Ejemplo de parámetros para un RIFS con transiciones restringidas.

| $w_i$ | $a_i$ | $b_i$ | $c_i$ | $d_i$ | $e_i$ | $f_i$ | $p_{i1}$ | $p_{i2}$ | $p_{i3}$ |
|-------|-------|-------|-------|-------|-------|-------|----------|----------|----------|
| 1     | 0.5   | 0     | 0     | 0.5   | 0     | 0     | 0.3      | 0.7      | 0        |
| 2     | 0.5   | 0     | 0     | 0.5   | 0     | 0.5   | 0        | 0.6      | 0.4      |
| 3     | 0.5   | 0     | 0     | 0.5   | 0.5   | 0.5   | 0.5      | 0        | 0.5      |

**Tabla 2.3:** Ejemplo de parámetros para un RIFS con todas las transiciones permitidas.

| $w_i$ | $a_i$ | $b_i$ | $c_i$ | $d_i$ | $e_i$ | $f_i$ | $p_{i1}$ | $p_{i2}$ | $p_{i3}$ |
|-------|-------|-------|-------|-------|-------|-------|----------|----------|----------|
| 1     | 0.5   | 0     | 0     | 0.5   | 0     | 0     | 0.1      | 0.3      | 0.6      |
| 2     | 0.5   | 0     | 0     | 0.5   | 0     | 0.5   | 0.1      | 0.5      | 0.4      |
| 3     | 0.5   | 0     | 0     | 0.5   | 0.5   | 0.5   | 0.2      | 0.3      | 0.5      |



**Figura 2.10:** Diagramas de transición para las cadenas de Markov asociadas a los RIFS de las Tablas 2.2 (a) y 2.3 (b).

En un RIFS, el “estado” del sistema se identifica con la última transformación aplicada. Así, si el sistema se encuentra en el estado  $i$ , la siguiente transformación se elige según la distribución  $\{p_{ij}\}_{j=1}^N$ . El RIFS posee un atractor único en el espacio de medidas, denominado *medida invariante recurrente*, que puede aproximarse mediante una versión generalizada del “juego del caos” o mediante un algoritmo iterativo de copiado.

### 2.9.1. Método del Juego del Caos para RIFS

Se describe a continuación un procedimiento iterativo para generar trayectorias que, en el límite, aproximan la medida invariante de un sistema de funciones iteradas recurrentes (RIFS). El algoritmo



se ilustra con los parámetros de las Tablas 2.2 y 2.3. La relación entre las órbitas producidas y la medida invariante es similar a la ya estudiada para IFS con probabilidades ordinarias.

**Paso 1.** Elegir un punto inicial  $(x_0, y_0) \in \mathbb{R}^2$ . Conviene que este punto se encuentre cerca del atractor esperado.

**Paso 2.** Elegir un estado inicial  $s_0 \in \{1, 2, 3\}$ . Cualquier estado es válido. La cadena de Markov definida por la matriz  $(p_{ij})$  posee una única distribución estacionaria  $(m_1, m_2, m_3)$  con  $m_i > 0$ , que satisface

$$\sum_{i=1}^3 m_i p_{ij} = m_j, \quad j = 1, 2, 3.$$

Si se conoce esta distribución, es aconsejable tomar como estado inicial aquel con mayor  $m_j$ .

**Paso 3.** Seleccionar  $s_1 \in \{1, 2, 3\}$  de acuerdo con las probabilidades  $p_{s_0 j}$ , es decir, la probabilidad de que  $s_1 = j$  es  $p_{s_0 j}$ .

**Paso 4.** Calcular  $(x_1, y_1) = w_{s_1}(x_0, y_0)$ .

**Paso 5.** Seleccionar  $s_2 \in \{1, 2, 3\}$  según las probabilidades  $p_{s_1 j}$ .

**Paso 6.** Calcular  $(x_2, y_2) = w_{s_2}(x_1, y_1)$ .

$\vdots$

**Paso (2n+1).** Elegir  $s_n \in \{1, 2, 3\}$  con probabilidad  $p_{s_{n-1} j}$ .

**Paso (2n+2).** Calcular  $(x_n, y_n) = w_{s_n}(x_{n-1}, y_{n-1})$ .

$\vdots$

Al finalizar, se obtiene una órbita finita

$$Q(\text{num}) = \{(x_n, y_n) : n = 1, 2, \dots, \text{num}\}.$$

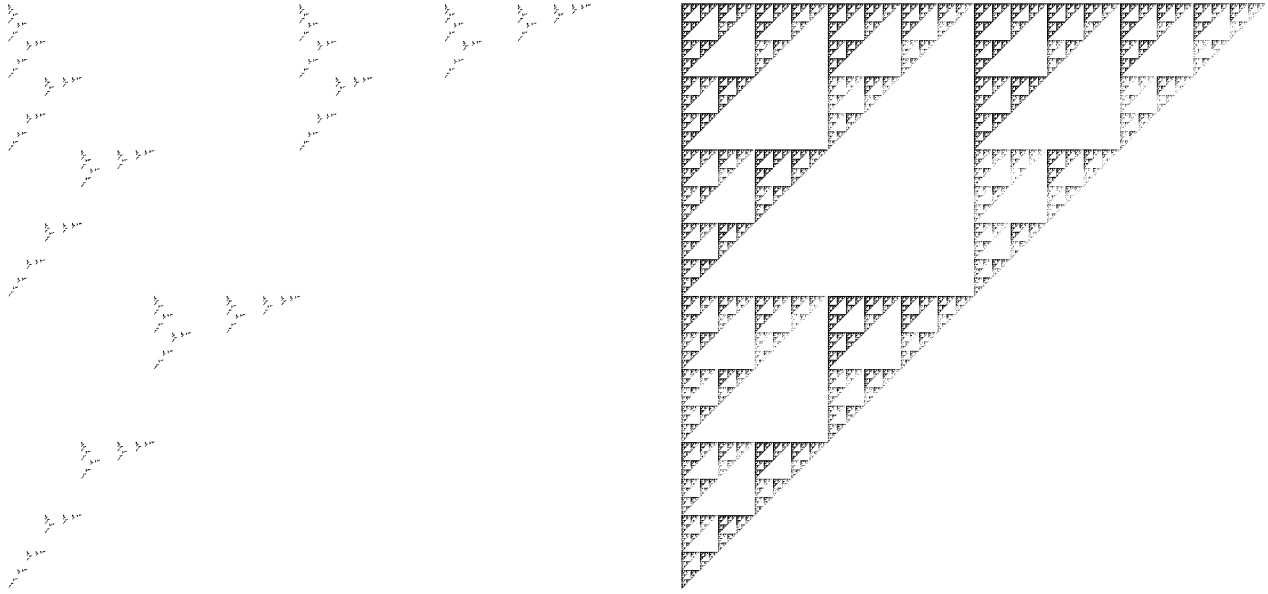
Con probabilidad uno, la medida invariante  $\mu \in \mathcal{P}([0, 1] \times [0, 1])$  del RIFS satisface

$$\lim_{\text{num} \rightarrow \infty} \mu(S) = \frac{\text{número de puntos en } S \cap Q(\text{num})}{\text{num}},$$

para todo conjunto medible  $S \subset \mathbb{R}^2$ .

La Figura 2.11 exhibe las aproximaciones obtenidas para los parámetros de las Tablas 2.2 y 2.3.

La Tabla 2.4 proporciona los coeficientes de un RIFS de cuatro aplicaciones definidas en el cuadrado unitario. La Figura 2.12 muestra el atractor resultante y la distribución de probabilidades en el espacio.



**Figura 2.11:** Aproximaciones a la medida invariante mediante el algoritmo del juego del caos para los RIFS de las Tablas 2.2 (izquierda) y 2.3 (derecha).

**Tabla 2.4:** Parámetros de un RIFS con cuatro transformaciones.

| $w_i$ | $a_i$ | $b_i$ | $c_i$ | $d_i$ | $e_i$ | $f_i$ | $p_{i1}$ | $p_{i2}$ | $p_{i3}$ | $p_{i4}$ |
|-------|-------|-------|-------|-------|-------|-------|----------|----------|----------|----------|
| 1     | 0.5   | 0     | 0     | 0.5   | 0     | 0     | 0.3      | 0.5      | 0.1      | 0.1      |
| 2     | 0.5   | 0     | 0     | 0.5   | 0     | 0.5   | 0.1      | 0.5      | 0        | 0.4      |
| 3     | 0.5   | 0     | 0     | 0.5   | 0.5   | 0.5   | 0.3      | 0.3      | 0.2      | 0.2      |
| 4     | 0.5   | 0     | 0     | 0.5   | 0.5   | 0     | 0.25     | 0.25     | 0.25     | 0.25     |

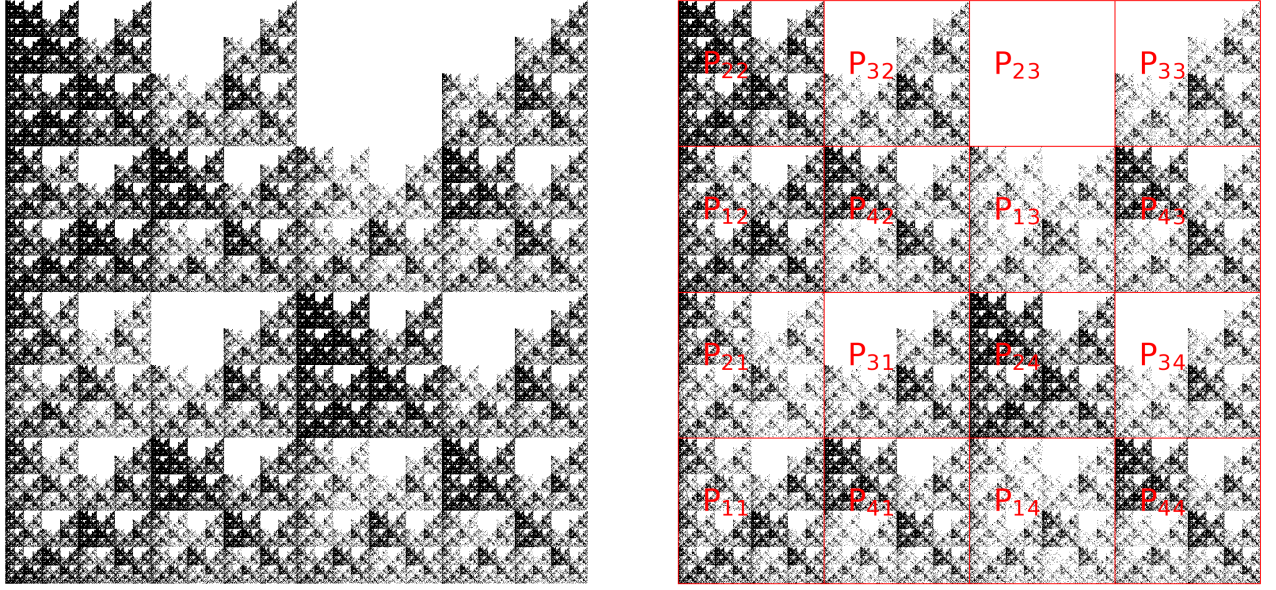
Se aprecia una zona en blanco en el atractor, correspondiente a la probabilidad de transición  $p_{23} = 0$ . Dado que ningún punto puede llegar a dicha región, su medida es nula. Este vacío se reproduce de manera autosimilar en todo el atractor.

Como ejemplo adicional, la Tabla 2.5 enumera seis configuraciones de probabilidades de transición para un RIFS con los mismos parámetros geométricos de la Tabla 2.4. Los atractores correspondientes se muestran en la Figura 2.13.

### 2.9.2. Método de la Máquina de Copiado

Este algoritmo, alternativo al anterior, también permite aproximar la medida invariante de un RIFS. La idea consiste en producir copias reducidas y filtradas de varias imágenes de entrada, que luego se combinan. La Figura 2.14 esquematiza el proceso:

1. Se toman tres imágenes de partida.
2. A cada una se le aplican las transformaciones  $w_1$ ,  $w_2$  y  $w_3$ . Las flechas en la Figura 2.14 indican



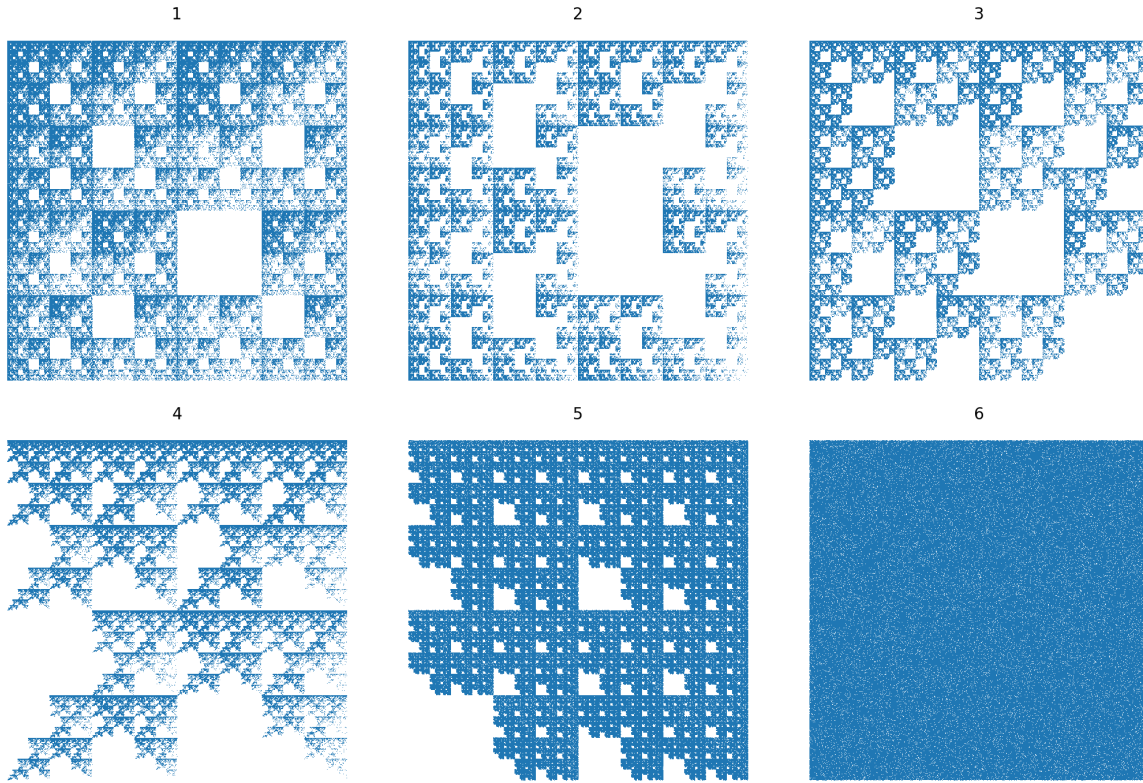
**Figura 2.12:** Atractor del RIFS definido por la Tabla 2.4. La imagen de la derecha ilustra la influencia de las probabilidades en diferentes regiones del cuadrado.

**Tabla 2.5:** Seis conjuntos de probabilidades de transición para un RIFS de cuatro transformaciones.

| No. | $w_i$ | $p_{i1}$ | $p_{i2}$ | $p_{i3}$ | $p_{i4}$ | No. | $w_i$ | $p_{i1}$ | $p_{i2}$ | $p_{i3}$ | $p_{i4}$ |
|-----|-------|----------|----------|----------|----------|-----|-------|----------|----------|----------|----------|
| 1.  | 1     | 0.3      | 0.5      | 0.1      | 0.1      | 2.  | 1     | 0.3      | 0.5      | 0.1      | 0.1      |
|     | 2     | 0.3      | 0.5      | 0.0      | 0        |     | 2     | 0.1      | 0.5      | 0        | 0.4      |
|     | 3     | 0.3      | 0.3      | 0.2      | 0.2      |     | 3     | 0.3      | 0.3      | 0.2      | 0.2      |
|     | 4     | 0.25     | 0.25     | 0.25     | 0.25     |     | 4     | 0.25     | 0.25     | 0.25     | 0.25     |
| 3.  | 1     | 0.3      | 0.5      | 0.1      | 0.1      | 4.  | 1     | 0.5      | 0.25     | 0        | 0.25     |
|     | 2     | 0.1      | 0.5      | 0        | 0.4      |     | 2     | 0.1      | 0.5      | 0.4      | 0        |
|     | 3     | 0.3      | 0.3      | 0.2      | 0.2      |     | 3     | 0.3      | 0.3      | 0.2      | 0.2      |
|     | 4     | 0.5      | 0        | 0.5      | 0        |     | 4     | 0.3      | 0.5      | 0.1      | 0.1      |
| 5.  | 1     | 0        | 0.3      | 0.3      | 0.4      | 6.  | 1     | 0.25     | 0.25     | 0.25     | 0.25     |
|     | 2     | 0.25     | 0.25     | 0.25     | 0.25     |     | 2     | 0.25     | 0.25     | 0.25     | 0.25     |
|     | 3     | 0.25     | 0.25     | 0.25     | 0.25     |     | 3     | 0.25     | 0.25     | 0.25     | 0.25     |
|     | 4     | 0.25     | 0.25     | 0.25     | 0.25     |     | 4     | 0.25     | 0.25     | 0.25     | 0.25     |

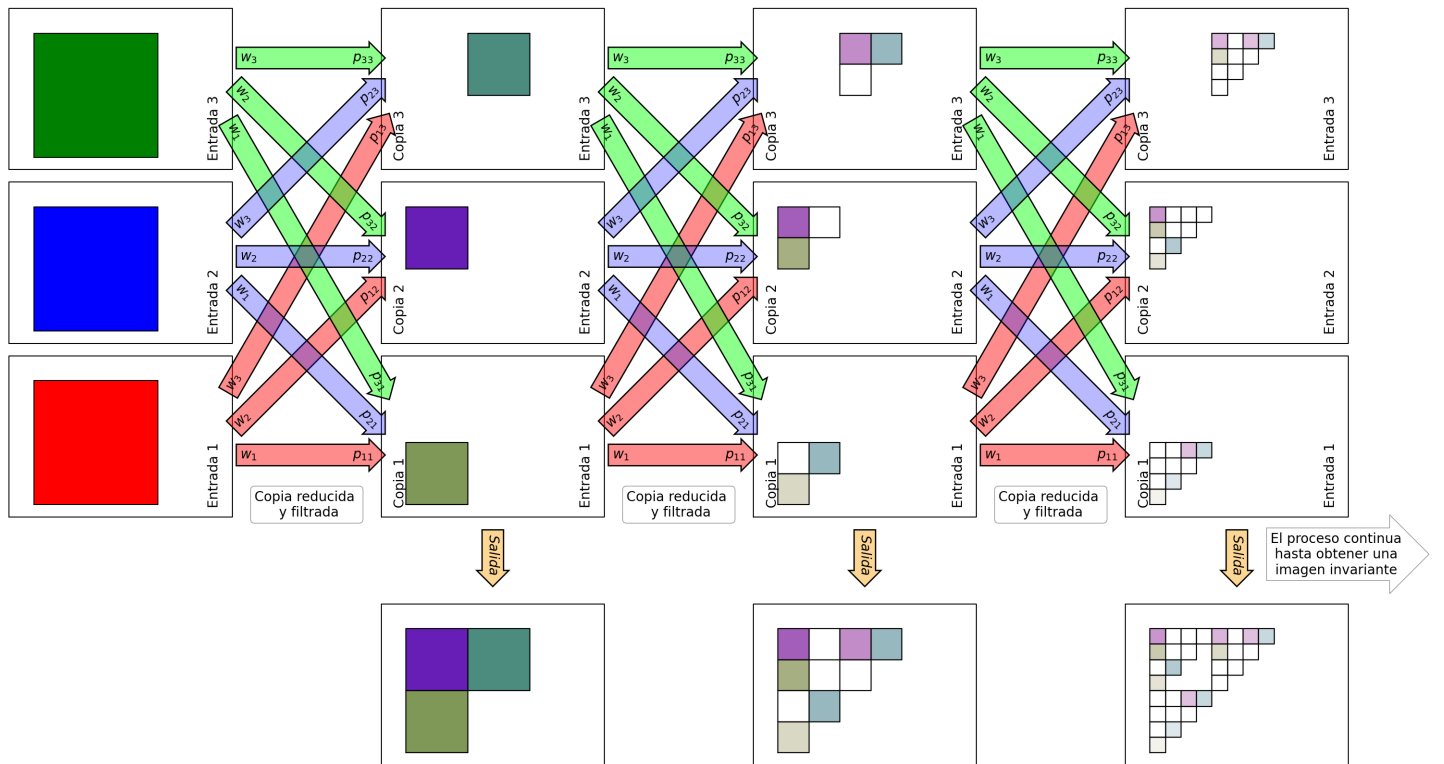
a qué entrada se asigna cada copia.

- Cada copia se atenúa mediante un factor  $p_{ij}$  que depende de la imagen de origen  $i$  y de la transformación  $w_j$  aplicada. Si  $p_{ij} = 0$ , la copia desaparece. En el ejemplo, la imagen 1 se filtra con factores  $p_{11}$ ,  $p_{12}$  y  $p_{13}$ .
- Las copias obtenidas se superponen para formar una nueva imagen.



**Figura 2.13:** Atractores generados por los seis conjuntos de probabilidades de la Tabla 2.5.

5. Se repite el proceso hasta que la imagen resultante no cambia apreciablemente.



**Figura 2.14:** Esquema del método de la máquina de copiado para un RIFS con tres estados. Se emplean los parámetros de la Tabla 2.2; el atractor obtenido coincide con el de la Figura 2.11.

La transparencia de cada copia se determina por una combinación lineal de las imágenes de entrada ponderadas por las probabilidades de transición. Por ejemplo, la copia destinada a la entrada 1 es

$$\text{copia}_1 = p_{11} \times \text{Entrada}_1 + p_{21} \times \text{Entrada}_2 + p_{31} \times \text{Entrada}_3.$$

Análogamente,

$$\text{copia}_2 = p_{12} \times \text{Entrada}_1 + p_{22} \times \text{Entrada}_2 + p_{32} \times \text{Entrada}_3,$$

$$\text{copia}_3 = p_{13} \times \text{Entrada}_1 + p_{23} \times \text{Entrada}_2 + p_{33} \times \text{Entrada}_3.$$

La imagen final (atractor) se obtiene superponiendo las tres copias:

$$\text{Atractor} = \text{copia}_1 + \text{copia}_2 + \text{copia}_3.$$

Los valores de transparencia así obtenidos corresponden a la medida invariante del sistema. Esta construcción se relaciona con el operador de Markov para RIFS, que se discutirá en la Sección sobre vectores de medida.

### 2.9.3. RIFS y Secuencias de ADN

Las representaciones gráficas de secuencias de ADN mediante el juego del caos han sido utilizadas para visualizar patrones en cadenas de nucleótidos [4]. Una secuencia de ADN es una cadena de símbolos en el alfabeto  $\{A, C, G, T\}$  (adenina, citosina, guanina, timina). Para traducirla a una imagen, se asocia cada letra a una esquina de un cuadrado. Se parte del centro del cuadrado; para cada letra de la secuencia, se traza un segmento desde la posición actual hacia la esquina correspondiente y se marca el punto medio. Repitiendo este procedimiento se genera una nube de puntos como la mostrada en la Figura 2.15 (izquierda), tomada de [4].

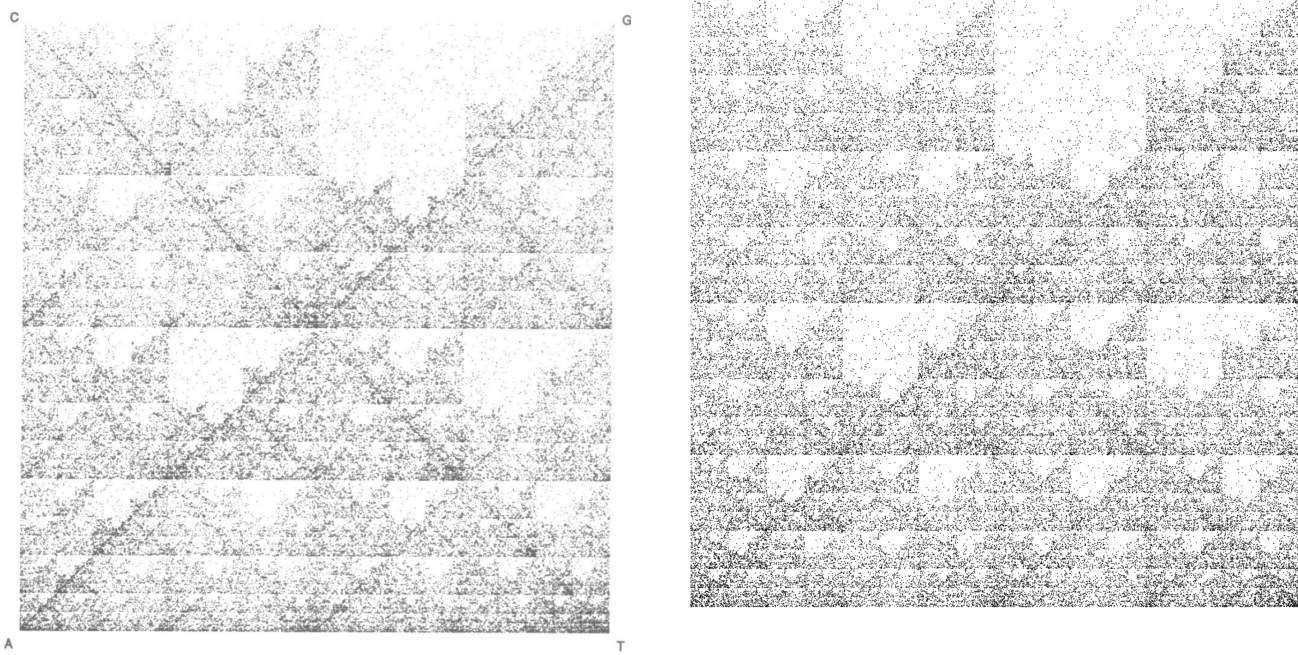
Un RIFS puede emplearse como modelo simplificado para simular dichas representaciones. La Tabla 2.7 contiene los parámetros de un RIFS cuyas probabilidades de transición se estiman a partir de las frecuencias de dinucleótidos en la secuencia. Concretamente, para cada par de nucleótidos  $X, Y \in \{A, C, G, T\}$ , la probabilidad  $P_{XY}$  se calcula como

$$P_{XY} = \frac{n_{XY}}{n_{XA} + n_{XC} + n_{XG} + n_{XT}},$$

donde  $n_{XY}$  es el número de veces que aparece el dinucleótido  $XY$  en la secuencia. Los valores  $n_{XY}$  para la región de beta globina humana se listan en la Tabla 2.6.

La Figura 2.15 (derecha) muestra la simulación obtenida con los parámetros de la Tabla 2.7. Se observa una notable similitud entre el patrón original y la simulación.

En la Figura 2.16 se presenta una simulación con mayor número de puntos, lo que permite apreciar con más detalle la estructura fractal del atractor.



**Figura 2.15:** Izquierda: representación de la secuencia de la región de beta globina humana mediante el juego del caos. Derecha: simulación mediante un RIFS con los parámetros de la Tabla 2.7.

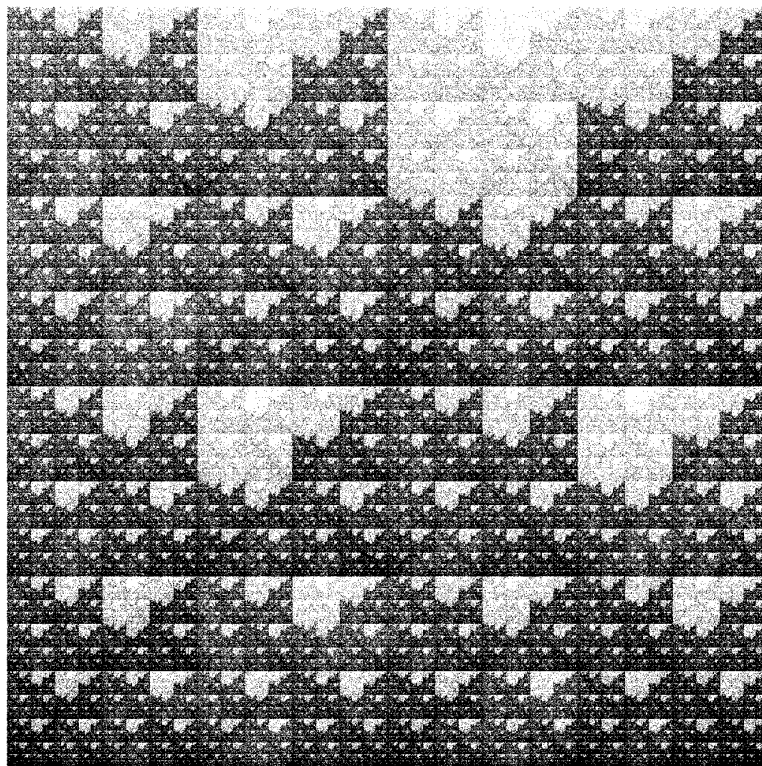
**Tabla 2.6:** Frecuencias de dinucleótidos  $n_{XY}$  en la secuencia de beta globina humana.

| $X \backslash Y$ | A    | C    | G    | T    | Total |
|------------------|------|------|------|------|-------|
| A                | 7239 | 3658 | 5227 | 5945 | 22069 |
| C                | 5166 | 3293 | 502  | 5207 | 14168 |
| G                | 4580 | 2839 | 3676 | 3694 | 14789 |
| T                | 5087 | 4379 | 5383 | 7444 | 22293 |

**Tabla 2.7:** Parámetros de un RIFS que simula la secuencia de ADN de beta globina humana.

| $w_i$ | $a_i$ | $b_i$ | $c_i$ | $d_i$ | $e_i$ | $f_i$ | $p_{i1}$ | $p_{i2}$ | $p_{i3}$ | $p_{i4}$ |
|-------|-------|-------|-------|-------|-------|-------|----------|----------|----------|----------|
| 1     | 0.5   | 0     | 0     | 0.5   | 0     | 0     | 0.328    | 0.166    | 0.237    | 0.269    |
| 2     | 0.5   | 0     | 0     | 0.5   | 0     | 0.5   | 0.365    | 0.233    | 0.035    | 0.367    |
| 3     | 0.5   | 0     | 0     | 0.5   | 0.5   | 0.5   | 0.309    | 0.192    | 0.249    | 0.250    |
| 4     | 0.5   | 0     | 0     | 0.5   | 0.5   | 0     | 0.228    | 0.197    | 0.241    | 0.334    |





**Figura 2.16:** Simulación con dos millones de puntos de la secuencia de beta globina humana mediante el RIFS de la Tabla 2.7.





# Capítulo 3

## Multifractales

Este capítulo aborda el marco teórico de la multifractalidad, presentándolo como una evolución y ampliación de los conceptos fundamentales de la geometría fractal. En contraste con los fractales clásicos, descritos por una dimensión única, los objetos multifractales exigen para su representación completa un espectro continuo de exponentes dimensionales que capturen las heterogeneidades en las propiedades de escalamiento de distribuciones complejas. Se analizan medidas ejemplares, entre ellas la ligada al conjunto de Cantor y las creadas a través de métodos iterativos de división binaria. Se introducen y discuten herramientas analíticas centrales como la dimensión de Rényi, el espectro multifractal  $f(\alpha)$  y las transformadas de Legendre. Los dominios de aplicación explorados comprenden desde fenómenos turbulentos y atractores caóticos en sistemas dinámicos hasta distribuciones multinomiales, revelando conexiones importantes con la termodinámica estadística y los principios de la teoría de la información. La presentación establece los fundamentos necesarios para el estudio de distribuciones irregulares en sistemas naturales y de elevada complejidad.

Imagine una magnitud física, como una densidad o concentración, distribuida de manera no homogénea dentro de un volumen de dimensión euclidiana  $L^E$ . Ejemplos típicos son la distribución espacial de la población en un continente, la concentración de elementos raros como el tántalo en la corteza terrestre, o los patrones de disipación de energía en un fluido turbulento. Estos elementos raros suelen hallarse en altas concentraciones en muy pocos lugares, mientras que en la gran mayoría de sitios aparecen solo como trazas. De forma análoga, la energía disipada en turbulencia y la distribución de impurezas sobre una superficie exhiben configuraciones espaciales muy desiguales y concentradas en regiones específicas.

La teoría multifractal se dedica precisamente al estudio de cómo se distribuyen estas cantidades físicas sobre soportes geométricos, que pueden variar desde planos euclidianos y esferas hasta conjuntos de naturaleza fractal. Los multifractales surgen como una generalización necesaria de los fractales simples, ya que para caracterizarlos adecuadamente se requiere un espectro completo de dimensiones en lugar de un único exponente. En términos conceptuales, un objeto multifractal puede entenderse

como una superposición entrelazada de múltiples subconjuntos, cada uno con características fractales propias. Esta idea pionera fue introducida por Mandelbrot [13], quien identificó un amplio rango de aplicaciones potenciales, dando pie a un extenso y fructífero campo de investigación tanto teórica como aplicada.

### 3.1. Medidas que Presentan Organización Multifractal

Esta sección examina dos casos ilustrativos en una dimensión, generados por sistemas iterativos con asignación de pesos probabilísticos, con el objetivo de demostrar cómo revelar la estructura multifractal que subyace en ellos.

#### 3.1.1. Medida Asociada al Conjunto de Cantor

El conjunto de Cantor se genera mediante un proceso recursivo. Se parte del intervalo cerrado  $[0, 1]$ , el cual se subdivide en tres segmentos de igual longitud. El tercio central se elimina, y el procedimiento se aplica de manera iterativa a los dos intervalos restantes. De forma más precisa:

$$\begin{aligned} I_0 &= [0, 1], \\ I_1 &= [0, \frac{1}{3}] \cup [\frac{2}{3}, 1], \\ I_2 &= [0, \frac{1}{9}] \cup [\frac{2}{9}, \frac{1}{3}] \cup [\frac{2}{3}, \frac{7}{9}] \cup [\frac{8}{9}, 1], \end{aligned}$$

y así sucesivamente.

El conjunto de Cantor se define como la intersección infinita  $\mathcal{C} = \bigcap_{n=0}^{\infty} I_n$ . Su dimensión fractal (de Hausdorff) es  $D = -\log 2 / \log 3 = \log_3 2$ . Cada conjunto  $I_n$  sirve como una cubierta de orden  $n$ . Dentro de cada cubierta hay  $2^n$  subintervalos de longitud  $3^{-n}$ , conocidos como *cilindros* de orden  $n$ . Se denotan por  $C_{z_1 z_2 \dots z_n}$ , donde cada  $z_i \in \{0, 1, 2\}$ . Por convención, el dígito 0 corresponde al subintervalo izquierdo, 1 al central y 2 al derecho. En la construcción estándar del conjunto de Cantor, solo se permiten los dígitos 0 y 2, excluyendo sistemáticamente el intervalo central asociado al dígito 1.

La cubierta inicial  $I_0 = [0, 1]$  contiene todos los números reales cuya expansión ternaria es  $0.z_1 z_2 z_3 \dots$  con  $z_i \in \{0, 1, 2\}$ . Si asignamos probabilidades fijas:  $p_0 \in [0, 1]$  al dígito 0,  $p_1 = 0$  al dígito 1, y  $p_2 = 1 - p_0$  al dígito 2, entonces el peso estadístico total satisface:

$$\mu(I_N) = \lim_{N \rightarrow \infty} (p_0 + p_1 + p_2)^N = \lim_{N \rightarrow \infty} (p_0 + p_2)^N = 1.$$

En el lenguaje de los sistemas iterativos de funciones (IFS) con pesos, el conjunto de Cantor se

describe como:

$$\left\{ X = [0, 1]; w_0(x) = \frac{x}{3}, w_1(x) = \frac{x}{3} + \frac{1}{3}, w_2(x) = \frac{x}{3} + \frac{2}{3}; p_0, p_1 = 0, p_2 \right\}.$$

Después de la primera iteración, a la porción izquierda  $C_0 = [0, 1/3]$  se le asigna una medida  $\mu_0 = p_0$ , y a la porción derecha  $C_2 = [2/3, 1]$  se le asigna  $\mu_2 = p_2$ , tal como se representa en la Figura 3.1. El segmento central  $C_1 = [1/3, 2/3]$  recibe medida nula ( $\mu_1 = p_1 = 0$ ). Los puntos en  $I_1$  tienen expansiones ternarias que comienzan con 0.0... (perteneciendo a  $C_0$ ) o con 0.2... (perteneciendo a  $C_2$ ), con pesos correspondientes:

$$\mu(C_0) = p_0, \quad \mu(C_2) = p_2, \quad \mu(C_1) = 0.$$

Iterando este proceso se obtienen distribuciones de probabilidad cada vez más refinadas, mostradas en la Figura 3.1.

Para la cubierta de segundo orden  $I_2$ , existen cuatro cilindros de longitud  $1/9$ :

$$\begin{aligned} C_{00} &= [0, 1/9], & C_{02} &= [2/9, 1/3], \\ C_{20} &= [2/3, 7/9], & C_{22} &= [8/9, 1], \end{aligned}$$

con medidas  $\mu(C_{00}) = p_0^2$ ,  $\mu(C_{02}) = p_0 p_2$ ,  $\mu(C_{20}) = p_2 p_0$ , y  $\mu(C_{22}) = p_2^2$ . La medida total se conserva:  $\mu(I_2) = p_0^2 + 2p_0 p_2 + p_2^2 = (p_0 + p_2)^2 = 1$ . En general, para un orden  $n$ :

$$\mu(I_n) = \sum_{z_1, \dots, z_n \in \{0, 2\}} \mu(C_{z_1 \dots z_n}) = \sum_{k=0}^n \binom{n}{k} p_0^k p_2^{n-k} = (p_0 + p_2)^n = 1.$$

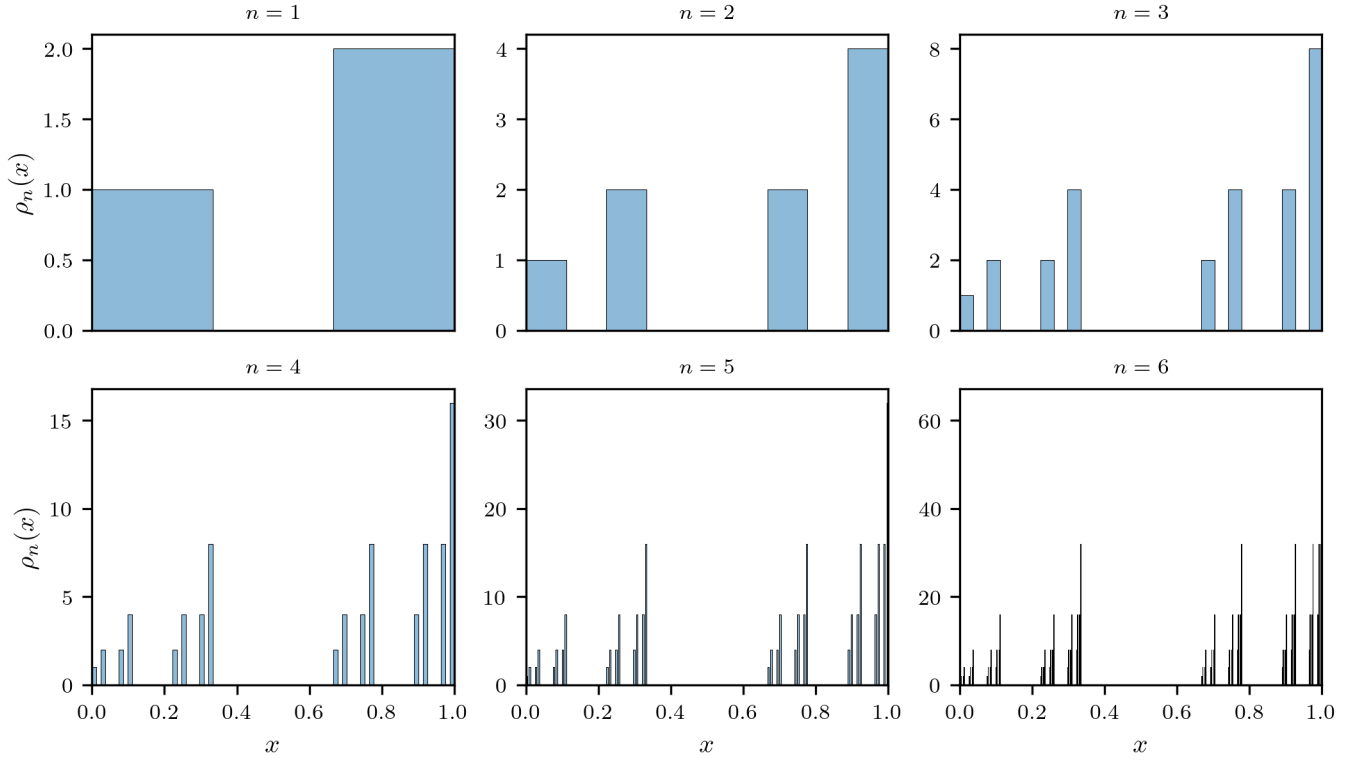
Considere el siguiente símil físico: una barra de densidad uniforme  $\rho_0 = 1$ , longitud  $l_0 = 1$  y masa total  $\mu_0 = 1$ . Si se corta en dos trozos de masa  $\mu_1 = 1/2$  y longitud  $l_1 = 1/3$ , la densidad en cada trozo se convierte en  $\rho_1 = \mu_1/l_1 = 3/2$ . Repitiendo este proceso  $n$  veces se obtienen  $2^n$  fragmentos, cada uno de longitud  $l_n = 3^{-n}$  y masa  $\mu_n = 2^{-n}$ . La masa total se mantiene constante:  $\sum_i \mu_i = 1$ . La masa de un fragmento escala con su longitud como  $\mu_i \sim l_i^\alpha$ , donde  $\alpha = \log 2 / \log 3$ . La densidad correspondiente,  $\rho_i = \mu_i / l_i^{\alpha-1}$ , aumenta a medida que  $l_i$  disminuye, lo que sugiere una divergencia en el límite  $l_i \rightarrow 0$ . El exponente  $\alpha$ , que gobierna esta singularidad, se denomina *exponente de singularidad* o *exponente de Hölder*.

Una medida  $\mu$  se considera singular en un entorno de  $x$  si, para una variable aleatoria  $\xi$ , satisface:

$$\mu(x \leq \xi \leq x + \Delta x) \sim (\Delta x)^{\alpha(x)}.$$

La densidad de probabilidad  $d\mu/dx$  existe únicamente cuando  $\alpha(x) = 1$  de manera global, pues:

$$\lim_{\Delta x \rightarrow 0} \frac{\mu(x \leq \xi \leq x + \Delta x)}{\Delta x} = \lim_{\Delta x \rightarrow 0} (\Delta x)^{\alpha(x)-1} = \begin{cases} 0, & \text{si } \alpha(x) > 1, \\ \infty, & \text{si } \alpha(x) < 1. \end{cases}$$



**Figura 3.1:** Distribuciones de probabilidad para las cubiertas de orden  $n = 1, 2, 3, 4, 5, 6$ , usando  $p_0 = 1/3$ ,  $p_2 = 2/3$ .

La masa acumulada hasta el punto  $x$  para la medida discreta  $\mu_n$  viene dada por:

$$C_n(x) = \int_0^x d\mu_n(t),$$

que en el límite  $n \rightarrow \infty$  converge a la llamada función de Cantor o “escalera del diablo”, una función de distribución acumulada singular. Si se considera una partición del intervalo en  $N$  segmentos  $J_i = [(i-1)/N, i/N]$ , se tiene que:

$$C_n(x) = \frac{1}{N} \sum_{i=1}^{k-1} y_i + y_k \left( x - \frac{k-1}{N} \right),$$

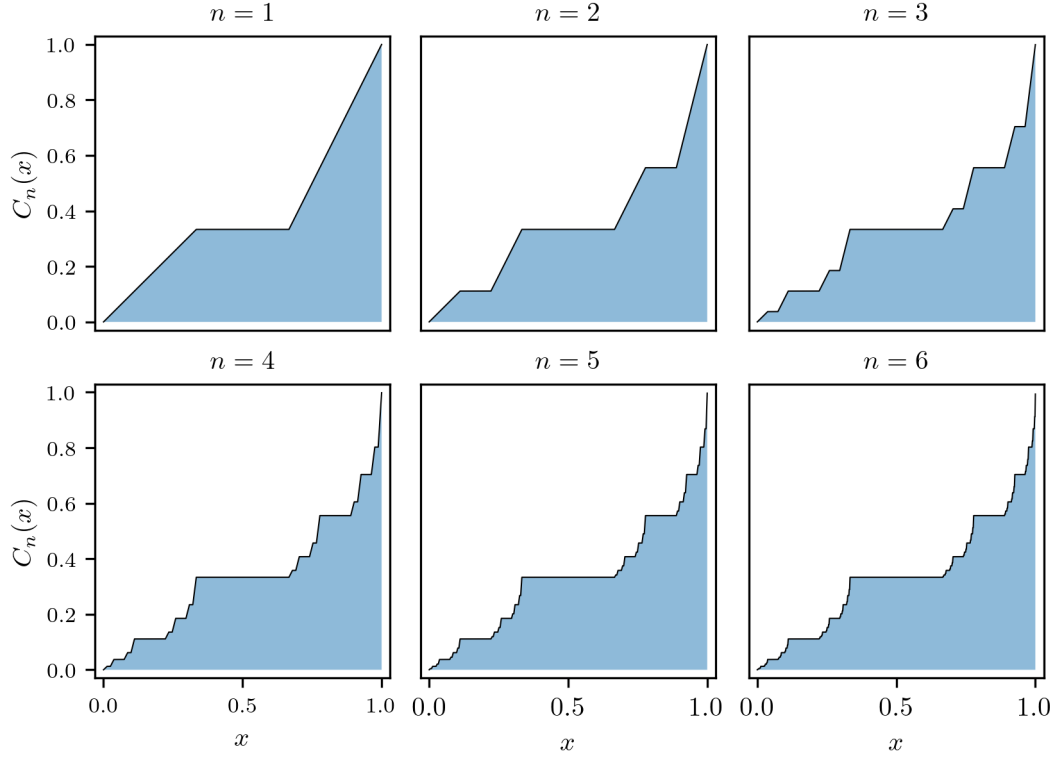
donde  $k$  es el índice tal que  $x \in J_k$ . En nuestro contexto, los cilindros  $C_{z_1 \dots z_n}$  juegan el papel de estos segmentos. La Figura 3.2 muestra las primeras iteraciones de esta función acumulativa, donde se aprecia su naturaleza autoafín.

Examinemos el comportamiento de escalamiento local. Para el primer cilindro (el más a la izquierda) en la etapa  $n$ , se cumple:

$$\frac{\log \mu_n([0, 3^{-n}])}{\log \delta_n} = \frac{\log(3^{-n})}{\log(3^{-n})} = 1,$$

donde  $\delta_n = 3^{-n}$  es la longitud del intervalo. Para el último cilindro (el más a la derecha), tenemos:

$$\frac{\log \mu_n([1 - 3^{-n}, 1])}{\log \delta_n} = \frac{\log((2/3)^n)}{\log(3^{-n})} = 1 - \log_3 2 \approx 0.3691.$$



**Figura 3.2:** Masa acumulada del conjunto de Cantor en función de la posición  $x$  en el intervalo  $[0, 1]$ .

Definamos  $I_n(\alpha)$  como la colección de subintervalos (cilindros) de orden  $n$  para los cuales el cociente anterior es igual a  $\alpha$ . Es decir, para  $C \in I_n(\alpha)$ ,

$$\frac{\log \mu(C)}{\log \delta_n} = \alpha.$$

Sea  $N_n(\alpha) = \#I_n(\alpha)$  el número de tales subintervalos. Por ejemplo, para  $n = 3$ :  $N_3(0.3691) = 1$ ,  $N_3(0.5794) = 3$ ,  $N_3(0.7897) = 3$ ,  $N_3(1) = 1$ . Se define la *dimensión prefractal* como:

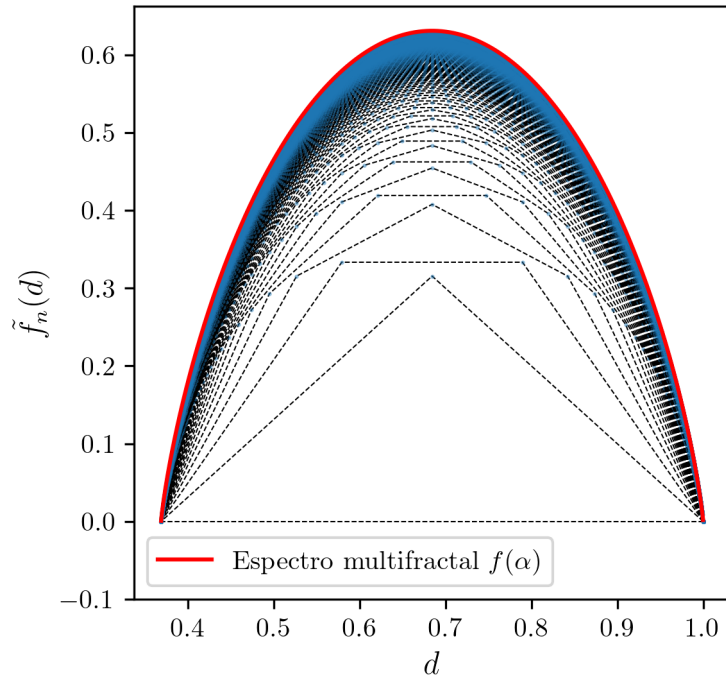
$$\tilde{f}_n(\alpha) = \frac{\log N_n(\alpha)}{\log(1/\delta_n)}.$$

La Figura 3.3 muestra  $\tilde{f}_n(\alpha)$  calculada para las primeras 500 iteraciones, junto con el espectro multifractal teórico  $f(\alpha) = \lim_{n \rightarrow \infty} \tilde{f}_n(\alpha)$ .

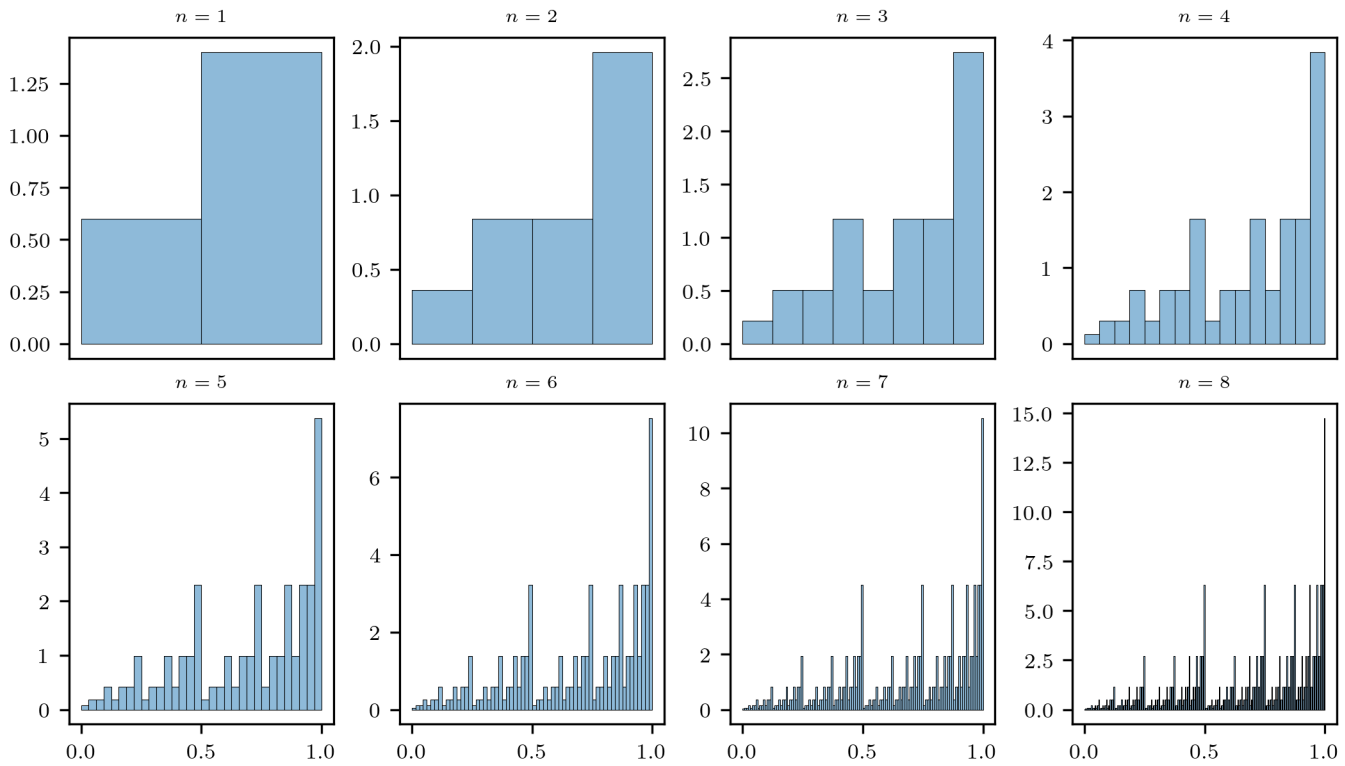
### 3.1.2. Fractales Autoafines de Soporte Continuo

Consideremos ahora un proceso iterativo diferente: la bipartición del intervalo unidad. En cada paso, el intervalo  $[0, 1]$  se divide en dos mitades iguales,  $C_0 = [0, 1/2]$  y  $C_1 = [1/2, 1]$ , a las cuales se asignan probabilidades  $p_0$  y  $p_1 = 1 - p_0$ , respectivamente. Este proceso se repite de forma indefinida, definiendo el sistema iterativo:

$$\left\{ X = [0, 1]; w_0(x) = \frac{x}{2}, w_1(x) = \frac{x}{2} + \frac{1}{2}; p_0, p_1 \right\}.$$



**Figura 3.3:** Espectro multifractal  $\tilde{f}(\alpha)$  para la medida del conjunto de Cantor. Los puntos representan los valores de  $\tilde{f}_n(\alpha)$  para las primeras 500 iteraciones.



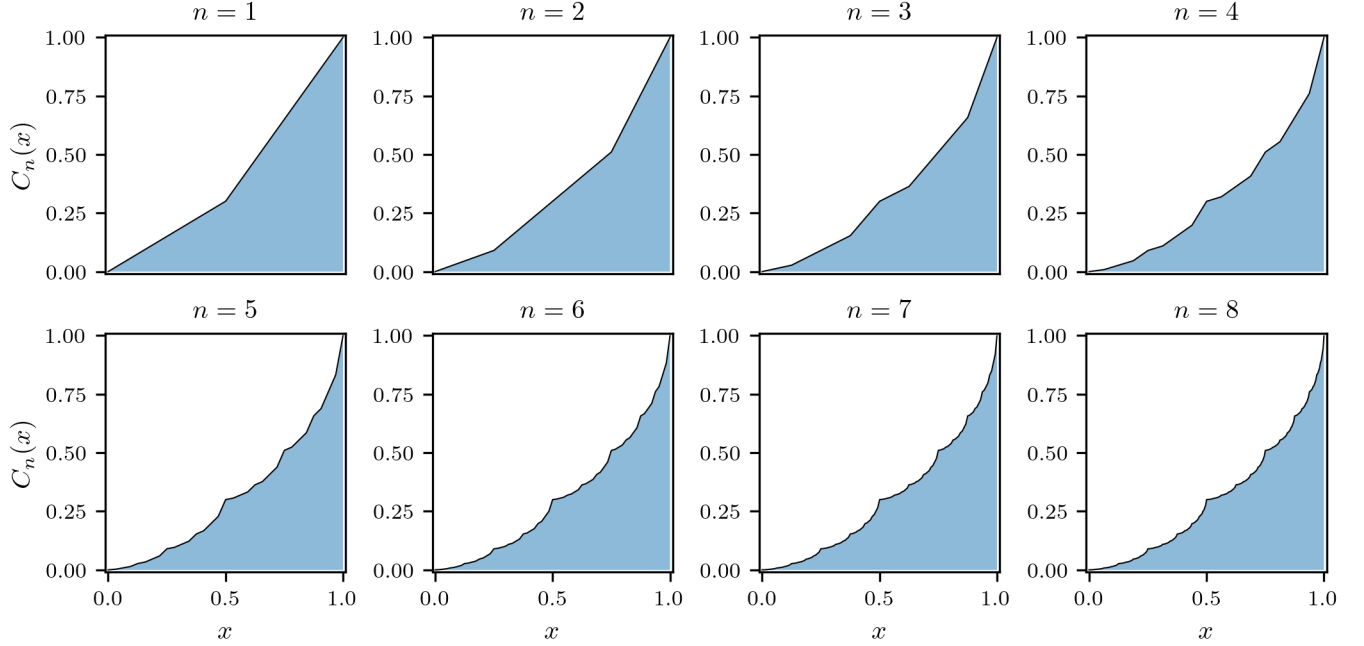
**Figura 3.4:** Distribución de probabilidad generada por bipartición iterativa. Muchas barras comparten el mismo valor de probabilidad.

La Figura 3.4 ilustra, en forma de histograma, la distribución de probabilidad resultante tras varias iteraciones.

Después de  $n$  iteraciones, el intervalo se divide en  $2^n$  segmentos de longitud  $\delta_n = 2^{-n}$ . La proba-

bilidad asociada al segmento que contiene a un punto  $x$  depende de la secuencia de dígitos binarios de  $x$ . En concreto, si entre los primeros  $n$  dígitos binarios de  $x$  hay  $k$  unos, entonces la probabilidad del segmento que contiene a  $x$  es  $p_0^k p_1^{n-k}$ . Por ejemplo, para  $n = 6$  y  $x = 1/5 = 0.00110011 \dots$ , los primeros seis dígitos contienen dos unos, por lo que la probabilidad en el segmento correspondiente (cilindro  $C_{001100}$ ) es  $p_0^2 p_1^4$ . Existen  $\binom{6}{2} = 15$  segmentos con esta misma probabilidad.

La correspondiente función de distribución acumulada, mostrada en la Figura 3.5, también exhibe propiedades de autoafinidad.



**Figura 3.5:** Función de distribución acumulada para el proceso de bipartición. Es una función continua y monótonamente creciente.

Una propiedad clave es la autoafinidad: la gráfica de la distribución en la mitad derecha del intervalo es una versión escalada verticalmente por un factor  $1/p_0$  de la gráfica en la mitad izquierda. En el límite  $n \rightarrow \infty$ , la función de distribución  $P(x)$  satisface la ecuación funcional:

$$P(x) = \frac{1}{p_0} P\left(\frac{x}{2}\right). \quad (3.1)$$

Esta relación se deduce de la observación de que la razón entre la probabilidad acumulada en la mitad izquierda y la acumulada en la mitad derecha es constante e igual a  $p_0/p_1$ . Dado que  $P(x/2) + [P(x) - P(x/2)] = P(x)$ , se obtiene fácilmente la ecuación anterior.

La dimensión fractal convencional del soporte (el intervalo completo) es 1, ya que  $N = 2^n$  segmentos de tamaño  $r = 2^{-n}$  cubren la totalidad del intervalo. La riqueza multifractal emerge al considerar no la geometría del soporte, sino la distribución espacial de la medida de probabilidad.

Después de  $n = 2m$  iteraciones, el número más abundante de segmentos son aquellos con probabilidad  $p_0^m p_1^m$  (es decir, con  $k = m$ ). Hay  $\binom{2m}{m}$  segmentos de este tipo. Estos segmentos están

localizados en posiciones cuya expansión binaria tiene exactamente  $m$  unos y  $m$  ceros en los primeros  $2m$  dígitos. La dimensión de Hausdorff de este conjunto de posiciones es:

$$D = \frac{1}{n} \log_2 \binom{n}{n/2}.$$

En general, hay  $\binom{n}{k}$  segmentos de longitud  $\delta_n = 2^{-n}$  con probabilidad  $p_0^k p_1^{n-k}$ . La probabilidad total contenida en todos los segmentos con un  $k$  dado es:

$$P_{k/n} = \binom{n}{k} p_0^k p_1^{n-k},$$

y la densidad correspondiente sería  $p_{k/n} = P_{k/n}/2^{-n}$ . Naturalmente,  $\sum_{k=0}^n P_{k/n} = 1$ .

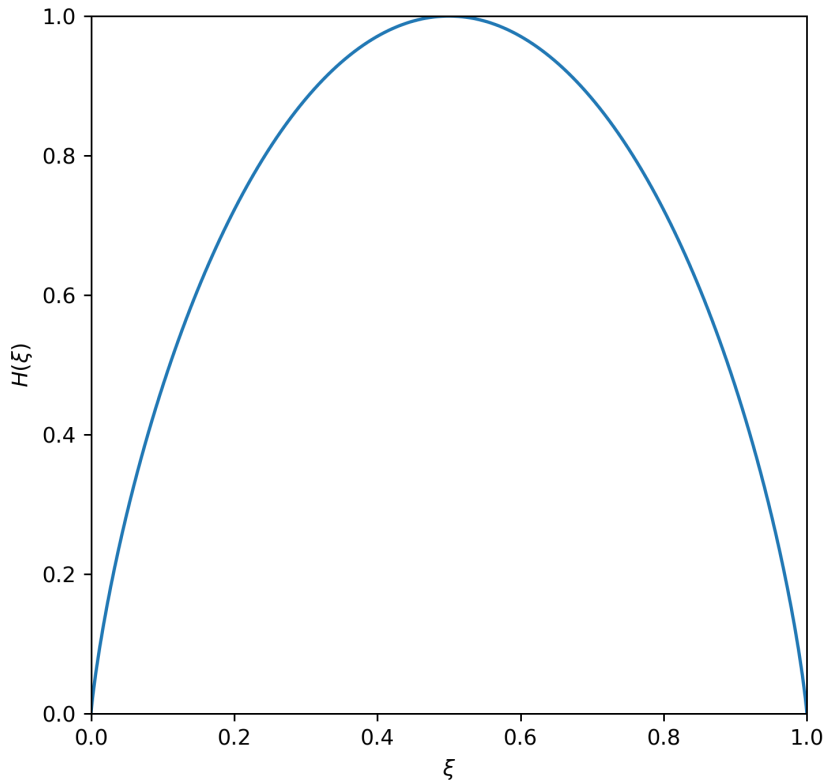
Para  $n$  grande podemos emplear la aproximación de Stirling, que conduce a:

$$\binom{n}{k} \approx 2^{nH(k/n)}, \quad (3.2)$$

donde  $H$  es la función de entropía binaria (o entropía de Shannon para dos estados):

$$H(\xi) = -\xi \log_2 \xi - (1 - \xi) \log_2 (1 - \xi). \quad (3.3)$$

La Figura 3.6 muestra la gráfica de esta función.



**Figura 3.6:** Gráfica de la función de entropía binaria  $H(\xi)$ .

La dimensión fractal del conjunto de puntos que pertenecen a segmentos con un exponente de singularidad dado (es decir, con una razón  $k/n$  fija) se obtiene como:

$$f(\xi) = \lim_{n \rightarrow \infty} \frac{\log \binom{n}{k}}{n \log(1/\delta_n)}, \quad \text{con } \xi = \frac{k}{n}.$$



Usando la aproximación (3.2) y notando que  $\log(1/\delta_n) = \log(2^n) = n \log 2$ , se obtiene:

$$f(\xi) = H(\xi). \quad (3.4)$$

Así, el espectro de dimensiones  $f(\xi)$  coincide con la función de entropía. Dependiendo del parámetro  $\xi$ , que está relacionado con la singularidad local de la medida, se obtienen diferentes dimensiones fractales. El rango de  $f(\xi)$  va desde 0 (en  $\xi = 0$  y  $\xi = 1$ ) hasta un máximo de 1 (en  $\xi = 1/2$ ). Los extremos  $\xi = 0, 1$  corresponden a conjuntos muy pequeños (un único segmento en cada extremo del intervalo), mientras que el valor  $\xi = 1/2$  corresponde al subconjunto más rico y extenso. Esta coexistencia de subconjuntos con diferentes dimensiones fractales, todos entrelazados en el mismo soporte, es la esencia del comportamiento multifractal.

### 3.1.3. La Dimensión de Rényi: Una Perspectiva desde la Teoría de la Información

El concepto de dimensión generalizada, introducido por Alfred Rényi en 1965, tiene sus raíces en la teoría de la información. En esta sección se presenta un breve recorrido histórico que conecta la entropía de Shannon con las dimensiones fractales, sin ánimo de ser exhaustivo. Para un tratamiento detallado puede consultarse la obra original de Rényi [42].

La teoría de la información, formalizada por Claude Shannon en 1948, se ocupa fundamentalmente de la cuantificación de la información. Un concepto básico es el de la información necesaria para especificar un elemento dentro de un conjunto finito. Si un conjunto  $E$  tiene  $n$  elementos y  $n = 2^N$ , entonces cada elemento puede identificarse de manera única mediante una cadena binaria de  $N$  bits. R. V. L. Hartley propuso que la información contenida en la especificación de un elemento es  $\log_2 n$  bits.

Por ejemplo, si  $n = 8$ , podemos etiquetar los elementos como 000, 001, 010, 011, 100, 101, 110, 111. La información requerida es  $\log_2 8 = 3$  bits.

Consideremos ahora un experimento aleatorio que clasifica los elementos de  $E$  en  $a$  categorías (subconjuntos)  $E_1, E_2, \dots, E_a$ . La probabilidad de que un elemento elegido al azar pertenezca a la categoría  $k$  es  $p_k$ , con  $\sum_{k=1}^a p_k = 1$ . Shannon definió la *entropía* de esta distribución como la información promedio obtenida al conocer la categoría de un elemento:

$$H_1(P) = - \sum_{k=1}^a p_k \log_2 p_k. \quad (3.5)$$

Rényi proporcionó una derivación notable de esta fórmula mediante el siguiente argumento probabilístico. Supongamos que hay un elemento particular (desconocido) en  $E$ . Realizamos una serie de experimentos independientes, en cada uno de los cuales todos los elementos se clasifican aleatoriamente en las  $a$  categorías según las probabilidades  $\{p_k\}$ . Después de cada experimento, observamos

en qué categoría cayó cada elemento. Tras  $m$  experimentos, cada elemento ha seguido una *trayectoria* de longitud  $m$  a través de las categorías. El elemento desconocido queda identificado de manera única cuando su trayectoria es diferente de la de todos los demás elementos.

Sea  $P_{nm}$  la probabilidad de que, después de  $m$  experimentos, el elemento desconocido tenga una trayectoria única. Rényi demostró que esta probabilidad tiene una expresión combinatoria compleja. Luego definió  $e_1(n, \epsilon)$  como el menor número de experimentos  $m$  necesario para que  $P_{nm} \geq 1 - \epsilon$ , con  $0 < \epsilon < 1$ . Su resultado fundamental establece que:

$$\lim_{n \rightarrow \infty} \frac{\log_2 n}{e_1(n, \epsilon)} = H_1(P), \quad (3.6)$$

donde  $H_1(P)$  es precisamente la entropía de Shannon (3.5). La interpretación es elegante:  $\log_2 n$  es la información total necesaria para identificar el elemento, y  $H_1(P)$  es la información promedio ganada por experimento. Así, el número típico de experimentos necesarios crece como  $\log_2 n / H_1(P)$ .

Rényi generalizó este concepto para definir una familia de entropías de orden  $q$ . Sea  $P_{nm}^q$  la probabilidad de que, tras  $m$  experimentos, ninguna trayectoria esté ocupada por  $q$  o más elementos. Definiendo  $e_q(n, \epsilon)$  de manera análoga, demostró que:

$$\lim_{n \rightarrow \infty} \frac{\log_2 n}{e_q(n, \epsilon)} = \left(1 - \frac{1}{q}\right) H_q(P),$$

donde la *entropía de Rényi* de orden  $q$  (con  $q \neq 1$ ) se define como:

$$H_q(P) = \frac{1}{1-q} \log_2 \left( \sum_{k=1}^a p_k^q \right). \quad (3.7)$$

Puede verificarse que  $\lim_{q \rightarrow 1} H_q(P) = H_1(P)$ , recuperando la entropía de Shannon.

Para conectar estas ideas con la noción de dimensión, consideremos una variable aleatoria  $X$  con distribución de probabilidad  $P = \{p_k\}$  sobre un conjunto numerable. Las entropías de Rényi cuantifican la información contenida en  $X$ . Si ahora consideramos una discretización del soporte de  $X$  con resolución  $\delta = 1/n$ , obtenemos una distribución discretizada  $P_n$ . La tasa a la que crece la entropía de Rényi cuando  $n$  aumenta (es decir, cuando la resolución se hace más fina) está relacionada con la dimensionalidad del soporte de la medida. Esto motiva la definición de la *dimensión de Rényi* de orden  $q$ :

$$D_q = \lim_{n \rightarrow \infty} \frac{H_q(P_n)}{\log_2 n}. \quad (3.8)$$

### Ejemplo: Conjunto de Cantor

Aplicemos esto a la medida del conjunto de Cantor. Después de  $n$  iteraciones, la variable discreta  $X_n$  toma valores en los  $2^n$  cilindros, con probabilidades  $\Pr(X_n = x_k) = p_0^{\# \text{ de ceros}} p_2^{\# \text{ de doses}}$ . La suma  $\sum_k p_k^q$  es fácil de calcular:

$$\sum_{k=1}^{2^n} [\Pr(X_n = x_k)]^q = \sum_{k=0}^n \binom{n}{k} p_0^{qk} p_2^{q(n-k)} = (p_0^q + p_2^q)^n.$$

Por lo tanto, la entropía de Rényi para la distribución  $P_n$  es:

$$H_q(P_n) = \frac{1}{1-q} \log_2((p_0^q + p_2^q)^n) = \frac{n}{1-q} \log_2(p_0^q + p_2^q).$$

La dimensión de Rényi resulta:

$$D_q = \lim_{n \rightarrow \infty} \frac{H_q(P_n)}{\log_2(3^n)} = \frac{1}{1-q} \cdot \frac{\log_2(p_0^q + p_2^q)}{\log_2 3} = \frac{1}{q-1} \log_3 \left( \frac{1}{p_0^q + p_2^q} \right).$$

Para el caso  $q = 1$ , aplicando la regla de L'Hôpital se obtiene la dimensión de información:

$$D_1 = -\frac{p_0 \log_3 p_0 + p_2 \log_3 p_2}{\log_3 3} = -(p_0 \log_3 p_0 + p_2 \log_3 p_2).$$

Mientras que la dimensión de Hausdorff del soporte (el conjunto de Cantor) es constante  $\log_3 2 \approx 0.6309$ , las dimensiones de Rényi  $D_q$  dependen de  $q$  y de los pesos  $p_0, p_2$ , caracterizando así cómo la medida se concentra o dispersa sobre el fractal. En las secciones siguientes se formalizarán estas definiciones en un contexto general.

### 3.1.4. Caracterización Multifractal mediante Dimensiones de Rényi

La descripción matemática de los multifractales puede realizarse mediante dos metodologías complementarias, como se detalla en [24]: el enfoque *reticular* y el *no reticular*. En el primer caso, se superpone al soporte del objeto una sucesión de mallas compuestas por celdas regulares de dimensión  $d$  cuyo lado  $\delta_n$  decrece gradualmente hacia cero a medida que  $n \rightarrow \infty$ . El segundo enfoque, en cambio, examina directamente la medida  $\mu$  en vecindades esféricas  $S_\delta(x)$  de radio  $\delta$  centradas en cada punto  $x$  del soporte. El análisis que se presenta a continuación se circunscribe exclusivamente a la metodología reticular, por su facilidad de implementación numérica y su claridad conceptual.

Imaginemos, pues, una malla regular que cubre por completo el soporte de la medida  $\mu$ . Esta malla está formada por celdas hipercúbicas  $d$ -dimensionales de arista  $\delta_n$ . Designaremos como  $B_{\delta_n}(k)$  a la celda  $k$ -ésima, donde el índice  $k$  recorre el conjunto  $I_n = \{k : \mu[B_{\delta_n}(k)] > 0\}$ , es decir, las celdas que contienen una parte no nula de la medida. El interés se centra en el comportamiento asintótico de ciertas cantidades cuando  $\delta_n \rightarrow 0$  (equivalente a  $n \rightarrow \infty$ ). Para cada número real  $q \in (-\infty, \infty)$ , se introduce la función auxiliar  $\tau(q)$  mediante el límite (si existe):

$$\tau(q) = \lim_{n \rightarrow \infty} \frac{\log \sum_{k \in K_n} \mu^q[B_{\delta_n}(k)]}{\log \delta_n}, \quad (3.9)$$

donde  $K_n$  denota un subconjunto apropiado de índices, usualmente todas las celdas con medida positiva. Esta función  $\tau(q)$  codifica información esencial sobre la homogeneidad escalante de la medida.

A partir de  $\tau(q)$  se define formalmente la familia de dimensiones generalizadas:

**Definición 3.1.1.** Para una medida  $\mu$ , la dimensión generalizada de Rényi (o dimensión de Rényi de orden  $q$ )  $D_q$  se define mediante las expresiones:

$$D_q = \begin{cases} \lim_{n \rightarrow \infty} \frac{\sum_{k \in K_n} \mu[B_{\delta_n}(k)] \log \mu[B_{\delta_n}(k)]}{\log \delta_n}, & \text{para } q = 1, \\ \frac{\tau(q)}{q-1}, & \text{para } q \neq 1, \end{cases}$$

asumiendo la existencia de los límites requeridos (para  $q = 1$  el límite del numerador, y para  $q \neq 1$  la existencia de  $\tau(q)$ ).

Una propiedad fundamental de estas dimensiones, válida para cualquier medida, fue establecida por Beck (1990):

**Teorema 3.1.1.** [Beck, 1990] Sean  $q$  y  $s$  números reales cualesquiera.

1. Si  $s > q$ , entonces se cumple  $D_s \leq D_q$ . Es decir, las dimensiones de órdenes mayores nunca exceden a las de órdenes menores.
2. Cuando  $s < q < 1$  o bien  $0 > s > q$ , se verifica la doble acotación:

$$\frac{s}{s-1} \frac{q-1}{q} \tilde{D}_q \leq D_s \leq D_q,$$

donde  $\tilde{D}_q$  representa una cota intermedia para la dimensión.

Basándonos en este teorema, podemos clasificar el grado de uniformidad de una medida. Se dice que  $\mu$  presenta *mínima uniformidad* si la cantidad  $\frac{q-1}{q} D_q$  se mantiene constante al variar  $q$ . Por el contrario, se alcanza la *máxima uniformidad* cuando todas las dimensiones  $D_q$  coinciden en un único valor constante, independiente de  $q$ .

### Ejemplo Concreto: La Medida de Cantor

Retomemos el ejemplo paradigmático de la medida sobre el conjunto de Cantor, con probabilidades  $p_0 = 1/3$  y  $p_2 = 2/3$  (recordando que  $p_1 = 0$ ). En este caso, un cálculo directo proporciona la forma explícita de la función  $\tau$ :  $\tau(q) = -\log_3(p_0^q + p_2^q)$ . La dimensión generalizada resulta entonces  $D(q) = \tau(q)/(q-1)$ . Se observa que  $D_q$  varía con  $q$ , lo que caracteriza un escenario de *no uniformidad*: la medida no se distribuye homogéneamente sobre el fractal.

Un subcaso especial de gran interés se produce cuando la medida se asigna de manera uniforme sobre el conjunto de Cantor, es decir, cuando  $p_0 = p_2 = 1/2$ . Entonces  $\tau(q) = (q-1) \log_3 2$ , lo que implica  $\tilde{D}(q) = \log_3 2$  para todo  $q$ . Aquí se manifiesta la *máxima uniformidad*, y el valor constante coincide, como era de esperar, con la dimensión fractal clásica (Hausdorff) del conjunto de Cantor.

Esta situación se generaliza a cualquier fractal autosimilar cuyas probabilidades asociadas a las contracciones sean idénticas ( $p_i = 1/N$  para  $i = 1, \dots, N$ ). En tal caso, se cumple que  $D_q = D_0$  para

todos los órdenes  $q$ . De hecho, no es necesario tomar el límite  $\delta_n \rightarrow 0$ ; una evaluación directa conduce a:

$$D_q = \frac{1}{q-1} \frac{\log(N(1/N)^q)}{\log \delta_n} = \frac{\log N}{\log(1/r)},$$

expresión que es manifiestamente independiente del parámetro  $q$ . Esto refleja la homogeneidad intrínseca de una medida distribuida por igual sobre todas las partes del fractal.

### 3.1.5. Interrelación entre $D_q$ , $\tau(q)$ y la Autosimilitud

La dimensión generalizada  $D_q$  no es más que una reparametrización del exponente  $\tau(q)$ , y su relación se despeja directamente de la definición cuando se considera un número fijo  $N$  de cajas de tamaño constante  $\delta_n = r$ . En el límite  $r \rightarrow 0$ , el valor de  $\tau(q)$  se mantiene, obteniéndose la conexión fundamental:

$$\tau(q) = (q-1)D_q. \quad (3.10)$$

Para un fractal autosimilar, donde conocemos tanto las probabilidades  $p_l$  como los factores de escala  $r_l$  asociados a cada contracción ( $l = 1, \dots, N$ ), la dimensión  $D_q$  puede determinarse de manera implícita combinando la relación anterior con la condición de normalización de la medida. Esto conduce a la ecuación:

$$\sum_{l=1}^N p_l^q r_l^{(q-1)D_q} = 1. \quad (3.11)$$

El caso especial  $q = 1$  merece atención particular, pues allí  $\tau(1) = 0$ . La dimensión  $D_1$  debe calcularse entonces mediante una derivada:

$$D_1 = \left. \frac{d\tau}{dq} \right|_{q=1}.$$

Al aplicar esta regla a la expresión general de  $\tau(q)$  se obtiene una fórmula explícita:

$$D_1 = - \frac{\sum_{l=1}^N p_l \log p_l}{\sum_{l=1}^N p_l \log r_l}.$$

Si todas las probabilidades son iguales ( $p_i = 1/N$ ), esta expresión se simplifica notablemente:

$$D_1 = \frac{N \log N}{\sum_{l=1}^N \log r_l}.$$

Alternativamente, tomando el límite  $q \rightarrow 1$  en la definición de  $D_q$ , se recupera una forma que involucra la entropía de Shannon  $H_1 = -\sum_l p_l \log p_l$ :

$$D_1 = \lim_{r \rightarrow 0} \frac{H_1}{\log r}.$$

La entropía  $H_1$  y la dimensión  $D_1$ , de profunda raigambre en la teoría de la información, desempeñan un papel crucial en el análisis de sistemas dinámicos no lineales. Específicamente, permiten

cuantificar la tasa a la que un sistema caótico genera información (o la pierde) a lo largo de su evolución temporal; en este contexto,  $H_1$  se denomina *entropía de Kolmogorov*.

Otro valor notable es el correspondiente a  $q = 2$ , que define la *dimensión de correlación*:

$$D_2 = \lim_{r \rightarrow 0} \frac{\log \left( \sum_{l=1}^N p_l^2 \right)}{\log r}.$$

Junto con  $D_0$  (dimensión de caja o de Hausdorff para el soporte) y  $D_1$ , la dimensión de correlación constituye una de las medidas fractales más utilizadas en aplicaciones prácticas debido a su relativa facilidad de estimación a partir de series temporales.

### 3.2. El Espectro Multifractal $f(\alpha)$ y su Dualidad con $\tau(q)$

Para una medida  $\mu$ , es posible definir un objeto geométrico que describe la distribución de las singularidades locales: el espectro multifractal  $f(\alpha)$ . Su definición formal es la siguiente.

**Definición 3.2.1.** Para cada exponente de singularidad  $\alpha > 0$  y cada tolerancia  $\epsilon > 0$ , consideremos el conjunto de cajas de la malla cuyo exponente de escalamiento local se aproxima a  $\alpha$ :

$$I_n(\alpha, \epsilon) = \left\{ k : \alpha - \epsilon < \frac{\log \mu[B_{\delta_n}(k)]}{\log \delta_n} \leq \alpha + \epsilon \right\}.$$

El espectro multifractal  $f(\alpha)$  se define mediante el límite doble (cuando existe):

$$f(\alpha) = \lim_{\epsilon \rightarrow 0} \lim_{n \rightarrow \infty} \frac{\log \#I_n(\alpha, \epsilon)}{-\log \delta_n},$$

donde  $\#I_n(\alpha, \epsilon)$  denota el número de cajas en el conjunto. Se admite  $f(\alpha) = -\infty$  cuando, para  $\epsilon$  suficientemente pequeño, el conjunto  $I_n(\alpha, \epsilon)$  está vacío para todo  $n$  grande.

Hemos presentado así dos descripciones alternativas para un multifractal: el espectro  $f(\alpha)$ , que asigna una dimensión fractal al conjunto de puntos con un exponente de Hölder local  $\alpha$  dado, y la familia de dimensiones generalizadas  $D_q$ , o equivalentemente la función  $\tau(q) = (1 - q)D_q$ . Ambas contienen la misma información geométrica sobre la medida, por lo que deben estar relacionadas matemáticamente. Efectivamente,  $f(\alpha)$  y  $\tau(q)$  son duales bajo una *transformada de Legendre*:

$$\tau(q) = \inf_{\alpha} \{ q\alpha - f(\alpha) \}, \quad (3.12)$$

$$f(\alpha) = \inf_q \{ q\alpha - \tau(q) \}. \quad (3.13)$$

De estas relaciones se deducen fórmulas diferenciales útiles:

$$\alpha = -\frac{d\tau}{dq}, \quad (3.14)$$

$$q = \frac{df}{d\alpha}. \quad (3.15)$$

La transformada de Legendre, herramienta central en mecánica clásica y termodinámica, establece un puente conceptual profundo entre la teoría multifractal y la mecánica estadística. La analogía es más que un mero cambio de variables: la ecuación que define  $\tau(q)$  juega un papel análogo al de la función de partición en física estadística. Bajo esta correspondencia, el exponente  $\alpha$  se identifica con la energía,  $\tau/q$  con la energía libre (de Helmholtz),  $f$  con la entropía y  $1/q$  con la temperatura. Esta conexión ha permitido importar técnicas y conceptos de la termodinámica al estudio de sistemas complejos.

### 3.3. Turbulencia

La turbulencia es un estado de flujo caótico y altamente irregular que se manifiesta en una enorme variedad de fenómenos naturales e industriales, desde la estela de un avión hasta los grandes flujos atmosféricos. Se caracteriza por la presencia de remolinos de múltiples escalas, fluctuaciones aleatorias en velocidad y presión, y una transferencia de energía no lineal desde escalas grandes hasta escalas pequeñas, donde finalmente se disipa por efectos viscosos.

Aunque las ecuaciones de Navier-Stokes gobiernan teóricamente el comportamiento de los fluidos, su extrema no linealidad y la complejidad de los flujos turbulentos han impulsado el desarrollo de modelos estadísticos y fenomenológicos. Uno de los logros de estos enfoques ha sido la caracterización de la disipación de energía cinética a escalas pequeñas, un proceso notablemente intermitente y espacialmente inhomogéneo.

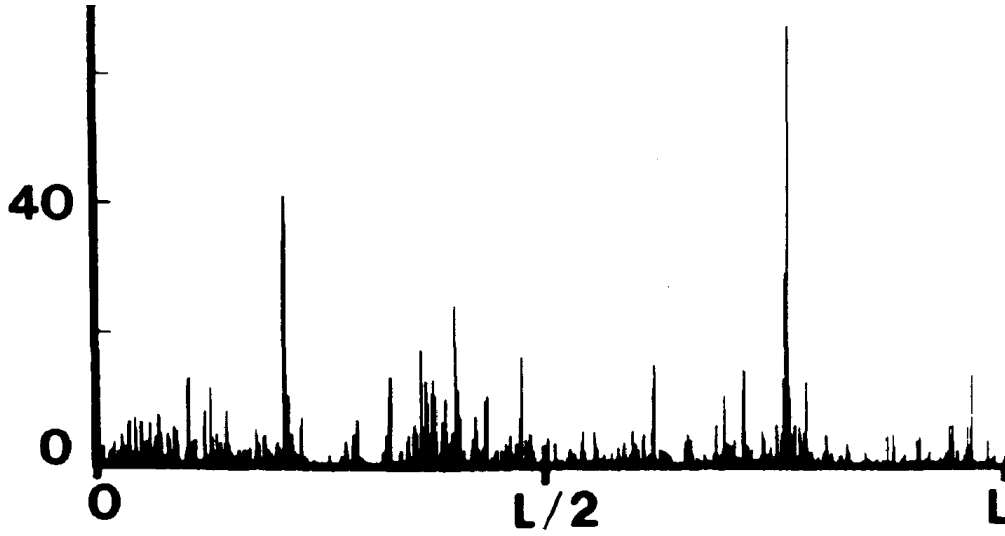
Mandelbrot (1974, 1975) fue pionero en sugerir que la turbulencia podría poseer una geometría fractal. Trabajos posteriores, como [5], han explorado esta hipótesis, determinando que es posible medir dimensiones fractales en estructuras turbulentas.

Una faceta particularmente interesante es la disipación de energía en turbulencia completamente desarrollada. Para explicar su intermitencia, se ha propuesto un *modelo de cascada multifractal* que se ajusta satisfactoriamente a datos experimentales (Figura 3.7).

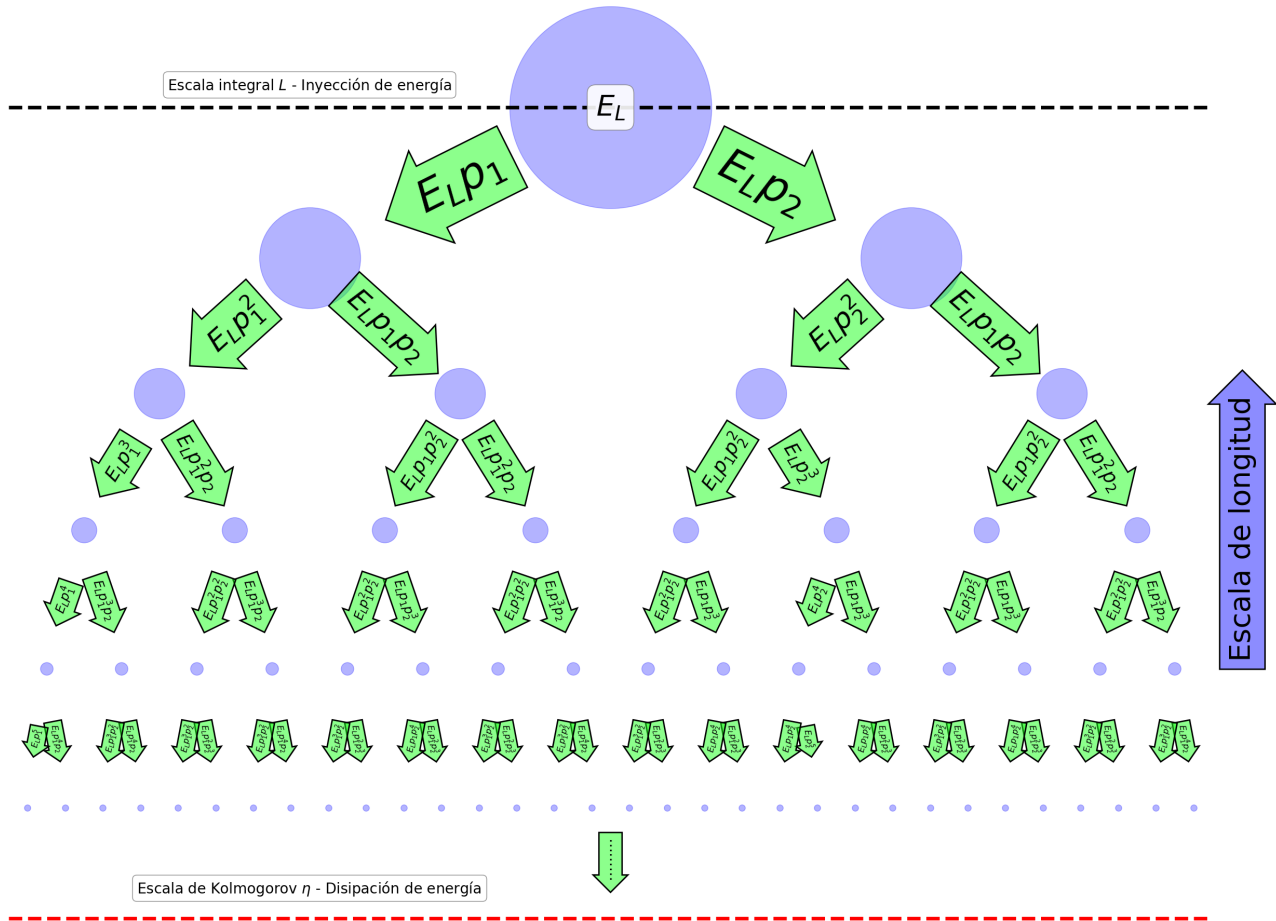
Este modelo, inspirado en la célebre cascada de Richardson (1922), conceptualiza la transferencia de energía como un proceso iterativo en el que un remolino grande se subdivide en remolinos más pequeños, transfiriendo su energía cinética en proporciones aleatorias. Estructuralmente, guarda una similitud notable con los esquemas iterativos de bipartición estudiados anteriormente.

En cada paso (ó “orden”)  $n$  de la cascada, una porción  $p_1$  de la energía se asigna aleatoriamente a uno de los dos sub-remolinos, y el complemento  $p_2 = 1 - p_1$  al otro. Tras  $n$  iteraciones, se genera una distribución espacial de energía disipada muy irregular, análoga a las distribuciones de probabilidad de los multifractales simples (Figura 3.9).

Para este modelo de cascada binaria, la dimensión generalizada  $D_q$  puede calcularse analítica-



**Figura 3.7:** Distribución espacial de la energía disipada normalizada,  $\epsilon/\epsilon_L$ , medida en un experimento de capa límite turbulenta. La intermitencia y las fuertes fluctuaciones son evidentes.

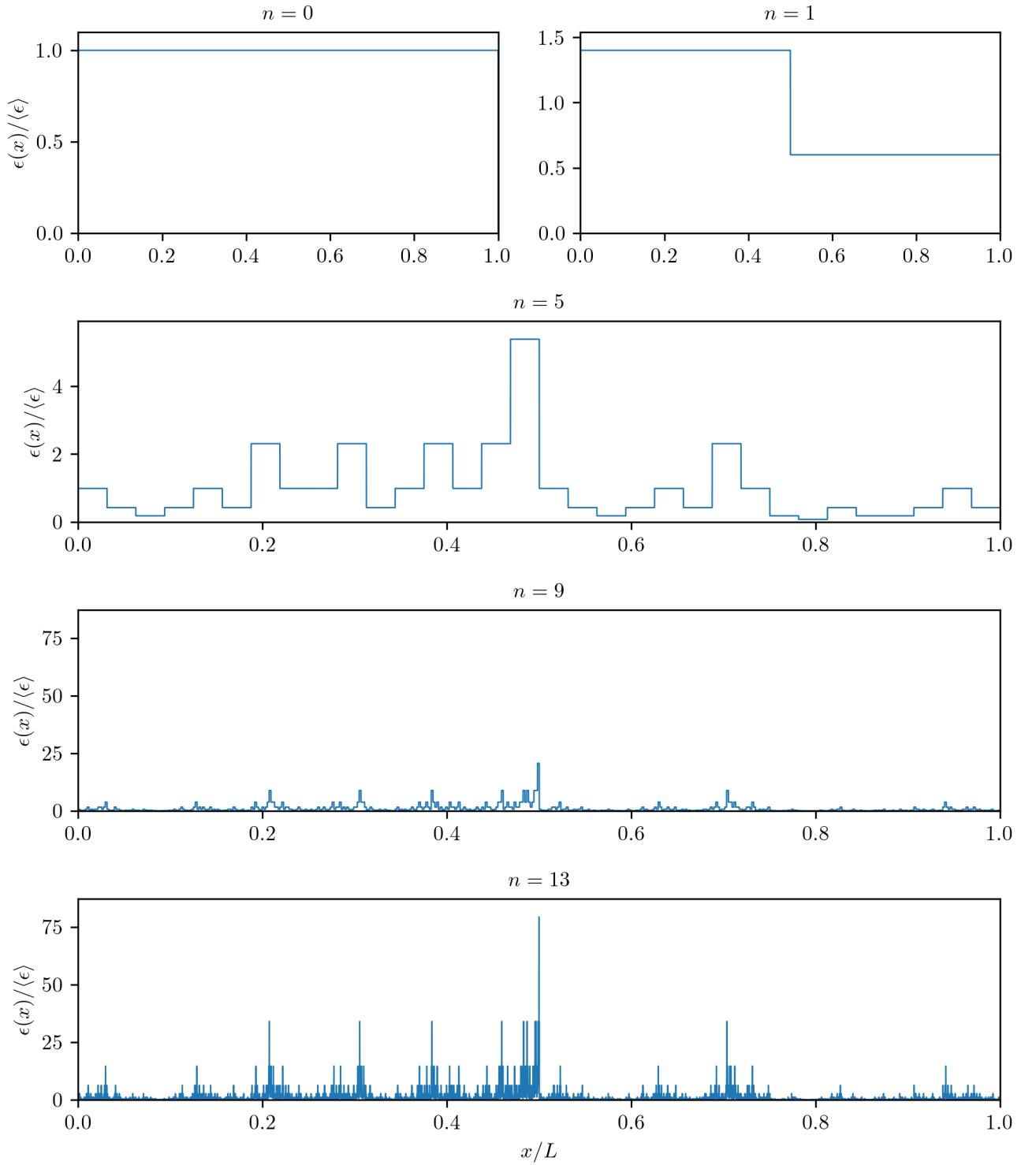


**Figura 3.8:** Representación esquemática de la cascada de energía de Richardson. Un remolino de escala integral  $L$  y energía  $E_L$  se divide en remolinos de tamaño  $L/2$ , que reciben fracciones aleatorias de la energía. El proceso continúa recursivamente hasta la escala de disipación  $\eta$ .

mente. Partiendo de la relación de escalamiento [15]:

$$\sum_i^N E_r^q = E_L^q \left( \frac{r}{L} \right)^{(q-1)D_q},$$





**Figura 3.9:** Evolución de la distribución de energía normalizada  $\epsilon/\langle\epsilon\rangle$  para los estados sucesivos  $n = 0, 1, 5, 9$  y  $12$ , utilizando  $p_1 = 0.7$  y  $p_2 = 0.3$ . El estado  $n = 12$  reproduce cualitativamente el patrón intermitente observado experimentalmente (Figura 3.7).

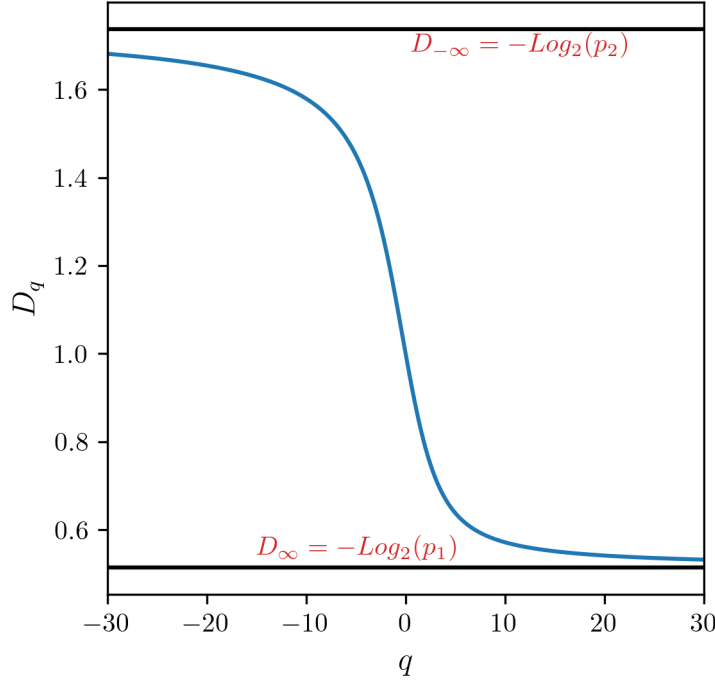
donde  $N = 2^n$ ,  $L$  es la escala integral,  $r = 2^{-n}L$  es la escala en el orden  $n$ , y  $E_r$  es la energía de un remolino individual, se llega tras un desarrollo algebraico a:

$$D_q = \frac{1}{1-q} \log_2(p_1^q + p_2^q).$$

Los límites para valores extremos de  $q$  son particularmente sencillos:

$$D_\infty = -\log_2 p_1, \quad D_{-\infty} = -\log_2 p_2 \quad (\text{suponiendo } p_1 > 1/2).$$

Estas expresiones permiten estimar el parámetro  $p_1$  a partir de datos experimentales [43]. Para  $p_1 = 0.7$ , el comportamiento de  $D_q$  se ilustra en la Figura 3.10, y el espectro multifractal asociado coincide con el calculado para otros modelos.



**Figura 3.10:** Curva teórica de la dimensión generalizada  $D_q$  para una sección unidimensional de un campo de disipación turbulento completamente desarrollado, según el modelo de cascada con  $p_1 = 0.7$ .

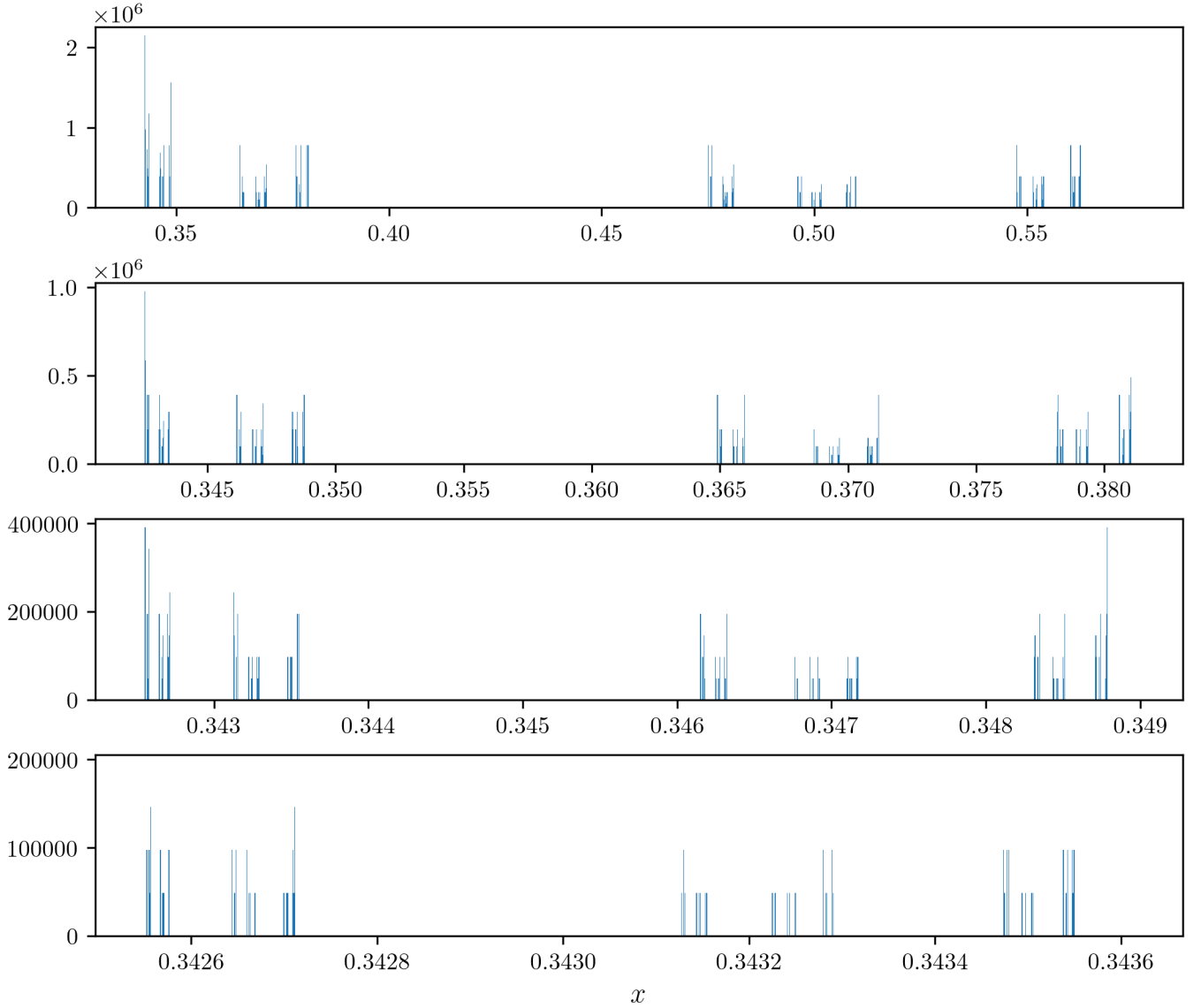
### 3.4. Atractores Extraños en Sistemas Dinámicos Caóticos

En el contexto de los sistemas dinámicos, un *atractor extraño* es un conjunto invariante hacia el cual convergen las órbitas de un sistema disipativo, y que exhibe las dos propiedades definitorias del caos: sensibilidad a las condiciones iniciales y una estructura fractal. A diferencia de los atractores clásicos (puntos fijos, ciclos límite), un atractor extraño no es una variedad diferenciable; su geometría es típicamente fractal y autosimilar en muchas escalas.

Un ejemplo canónico es el atractor del mapa logístico  $L_\lambda(x) = \lambda x(1 - x)$  en el régimen caótico. Para el valor crítico  $\lambda_\infty \approx 3.5699456 \dots$ , se sabe que la órbita genérica converge hacia un conjunto de tipo Cantor, un atractor extraño de medida nula dentro del intervalo  $[0, 1]$ .

Al simular una órbita larga (por ejemplo,  $10^8$  iteraciones) para  $\lambda = \lambda_\infty$  y construir histogramas en ventanas cada vez más pequeñas (Figura 3.11), se observa una estructura recurrente: ciertos

subintervalos aparecen poblados mientras otros permanecen vacíos, y este patrón se repite a diferentes escalas, sugiriendo fuertemente una autosimilitud estadística.



**Figura 3.11:** Histogramas de una órbita del mapa logístico para  $\lambda_\infty = 3.5699456\dots$ . La ventana superior muestra la distribución en  $[0.34, 0.58]$ ; las ventanas inferiores son ampliaciones sucesivas (escala reducida por factor 0.166), revelando una estructura fragmentada y auto-semejante.

Este atractor puede modelarse de manera aproximada mediante un sistema iterativo de funciones (IFS) con dos contracciones de factores de escala diferentes,  $r_1 = 0.408$  y  $r_2 = r_1^2 \approx 0.166$ , y probabilidades iguales  $p_1 = p_2 = 0.5$ . La dimensión fractal  $D_0$  se obtiene resolviendo la ecuación característica:

$$r_1^D + r_2^D = 1.$$

Con  $r_2 = r_1^2$ , esto se reduce a  $x + x^2 = 1$  con  $x = r_1^D$ , cuya solución positiva  $x = (\sqrt{5} - 1)/2 \approx 0.618$  conduce a  $D \approx 0.537$ .

Para este IFS, la función  $\tau(q)$  puede calcularse explícitamente:

$$\tau(q) = -\frac{\log \left[ \frac{1}{2} (\sqrt{1 + 2^{q+2}} - 1) \right]}{\log r_1}. \quad (3.16)$$

La dimensión generalizada se deriva entonces como  $D_q = \tau(q)/(q - 1)$  para  $q \neq 1$ . En los límites  $q \rightarrow \pm\infty$  se obtienen los valores extremos:

$$D_\infty = -(\log_2 r_1^2)^{-1} \approx 0.387, \quad D_{-\infty} = -(\log_2 r_1)^{-1} \approx 0.773.$$

El espectro multifractal  $f(\alpha)$  correspondiente puede obtenerse paramétricamente. Para una fracción  $\xi = k/n$  de pasos que eligen la contracción de factor  $r_1$ , el exponente de Hölder es:

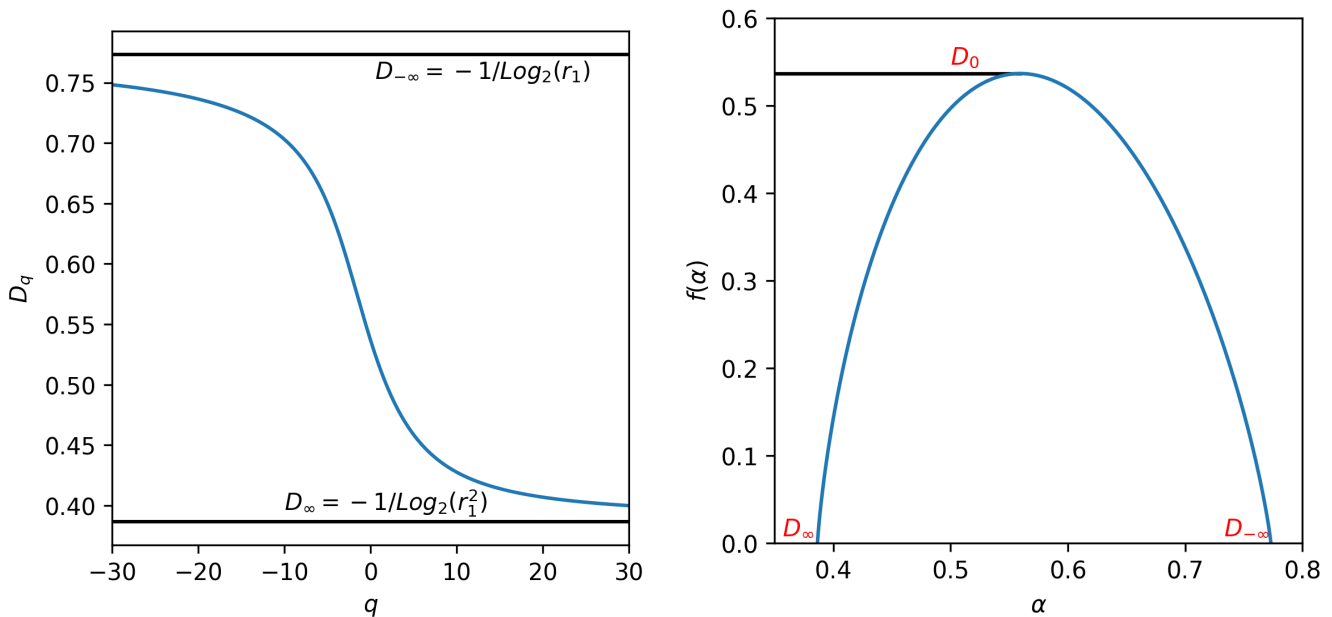
$$\alpha(\xi) = -\frac{1}{(2 - \xi) \log_2 r_1},$$

y la dimensión del subconjunto con ese exponente es:

$$f(\xi) = \frac{(1 - \xi) \log(1 - \xi) + \xi \log \xi}{(2 - \xi) \log r_1}. \quad (3.17)$$

Eliminando el parámetro  $\xi$  se obtiene  $f$  como función de  $\alpha$ .

La elección específica  $r_1 = 0.408$  modela el atractor extraño del mapa logístico con una discrepancia de sólo un 2.5 % respecto a estimaciones numéricas, indicando que, aunque no es perfectamente autosimilar, el modelo captura la esencia de su estructura multifractal. La Figura 3.12 muestra las curvas de  $D_q$  y  $f(\alpha)$  para este sistema.



**Figura 3.12:** (Izquierda) Dimensión generalizada  $D_q$  y (derecha) espectro multifractal  $f(\alpha)$  para el modelo IFS que aproxima el atractor extraño del mapa logístico en  $\lambda_\infty$ .

### 3.5. La Medida Multinomial: Un Marco General

Los ejemplos anteriores (Cantor, bipartición) son casos particulares de una construcción más general y poderosa: la *medida multinomial*. Esta medida se define sobre el intervalo unitario  $[0, 1]$  mediante un proceso iterativo de división y asignación de masa que generaliza naturalmente los esquemas binarios o ternarios.

El proceso inicia dividiendo el intervalo en  $b$  subintervalos de igual longitud ( $b \geq 2$ ). A cada subintervalo se le asigna una probabilidad  $p_z$  (con  $z = 0, 1, \dots, b-1$ ), tal que  $\sum_z p_z = 1$ . Luego, cada subintervalo se subdivide a su vez en  $b$  partes, redistribuyendo su masa según las mismas probabilidades  $p_z$ . Este procedimiento se repite ad infinitum.

Tras  $n$  iteraciones, el intervalo queda partido en  $b^n$  subintervalos de longitud  $\delta_n = b^{-n}$ , llamados *cilindros de orden  $n$* . Cada cilindro corresponde unívocamente a una secuencia finita  $\omega_n = z_1 z_2 \dots z_n$  de dígitos en base  $b$ , y su medida es el producto de las probabilidades asociadas a cada dígito:

$$\mu[C_{\omega_n}] = p_{z_1} p_{z_2} \dots p_{z_n} = p_0^{n\Phi_0(\omega_n)} p_1^{n\Phi_1(\omega_n)} \dots p_{b-1}^{n\Phi_{b-1}(\omega_n)},$$

donde  $\Phi_i(\omega_n)$  es la fracción de dígitos igual a  $i$  en la secuencia  $\omega_n$ .

Para excluir dígitos con probabilidad nula, se define el alfabeto activo  $\Omega = \{z : p_z > 0\}$  de tamaño  $s = \#\Omega$ . En el conjunto de Cantor estándar, por ejemplo,  $b = 3$  pero  $\Omega = \{0, 2\}$  pues  $p_1 = 0$ , luego  $s = 2$ .

#### 3.5.1. Teorema de Eggleston y Dimensiones Fractales

Una pregunta natural es: ¿cuál es la dimensión fractal del conjunto de puntos cuyas expansiones en base  $b$  tienen frecuencias asintóticas de dígitos dadas por un vector  $\phi = (\phi_0, \dots, \phi_{b-1})$ ? Este conjunto, denotado  $M(\phi)$ , está formado por la unión de todos los cilindros cuyas frecuencias empíricas tienden a  $\phi$ .

El célebre *teorema de Eggleston* proporciona la respuesta: la dimensión de Hausdorff de  $M(\phi)$  es

$$D_H(M(\phi)) = - \sum_{i \in \Omega} \phi_i \log_b \phi_i, \quad (3.18)$$

conviniendo que  $0 \log_b 0 = 0$ . Este resultado se demuestra utilizando la fórmula de Stirling para estimar el número de cilindros con frecuencias cercanas a  $\phi$ .

El intervalo unitario, bajo la medida multinomial, se revela así como un verdadero multifractal: es la unión de una infinidad de conjuntos  $M(\phi)$ , cada uno con una dimensión fractal dada por (3.18). La dimensión global del intervalo es 1, que corresponde al máximo de  $D_H(\phi)$ , alcanzado cuando todas las frecuencias son iguales ( $\phi_i = 1/b$ ), es decir, para los números *normales* en base  $b$ .

### Ejemplos Específicos

Para el caso binario ( $b = 2$ ), el teorema de Eggleston se reduce a la expresión ya vista para el espectro de la bipartición:  $f(\xi) = -[\xi \log_2 \xi + (1 - \xi) \log_2(1 - \xi)]$ .

Para el conjunto de Cantor ( $b = 3$  con  $p_1 = 0$ ), el espacio de frecuencias relevantes es bidimensional ( $\phi_0, \phi_2$  con  $\phi_0 + \phi_2 = 1$ ), y su dimensión está escalada por el factor  $\log 2 / \log 3$ :

$$D_H^C(\phi_0, \phi_2) = \frac{\log 2}{\log 3} (-\phi_0 \log_3 \phi_0 - \phi_2 \log_3 \phi_2).$$

Cuando  $\phi_0 = \phi_2 = 1/2$ , se recupera la dimensión máxima  $\log_3 2$ .

#### 3.5.2. Visualización: Espectros Ópticos Multifractales

Una forma intuitiva de apreciar la riqueza multifractal es mediante *espectros ópticos*. Para  $b = 3$ , cada conjunto  $M(\phi_0, \phi_1, \phi_2)$  puede representarse con un color que codifique su dimensión  $D_H$ . Con  $n = 8$  dígitos, existen 45 de estos conjuntos (Figura 3.13), dispersos de manera no contigua por el intervalo.

La superposición de todos ellos produce un patrón de bandas de colores, el espectro óptico del multifractal (Figura 3.14). Notablemente, si tomamos un solo cilindro de orden 8 y aplicamos el proceso de división multinomial dentro de él, recuperamos una estructura idéntica en miniatura (Figura 3.15), evidenciando la autosimilitud estadística del objeto.

La dimensión  $D_H(\phi_0, \phi_1, \phi_2)$  para  $b = 3$  define una superficie sobre el triángulo de posibles frecuencias (Figura 3.16), cuyo máximo está en el centro ( $\phi_i = 1/3$ ) y vale 1.

#### 3.5.3. Comportamiento Local y Exponente de Hölder

Dado que la medida multinomial es singular (no posee densidad), es instructivo estudiar su comportamiento local mediante el *exponente de Hölder*  $\alpha$ . Para un cilindro  $C_{\omega_n}$ , este exponente se define como:

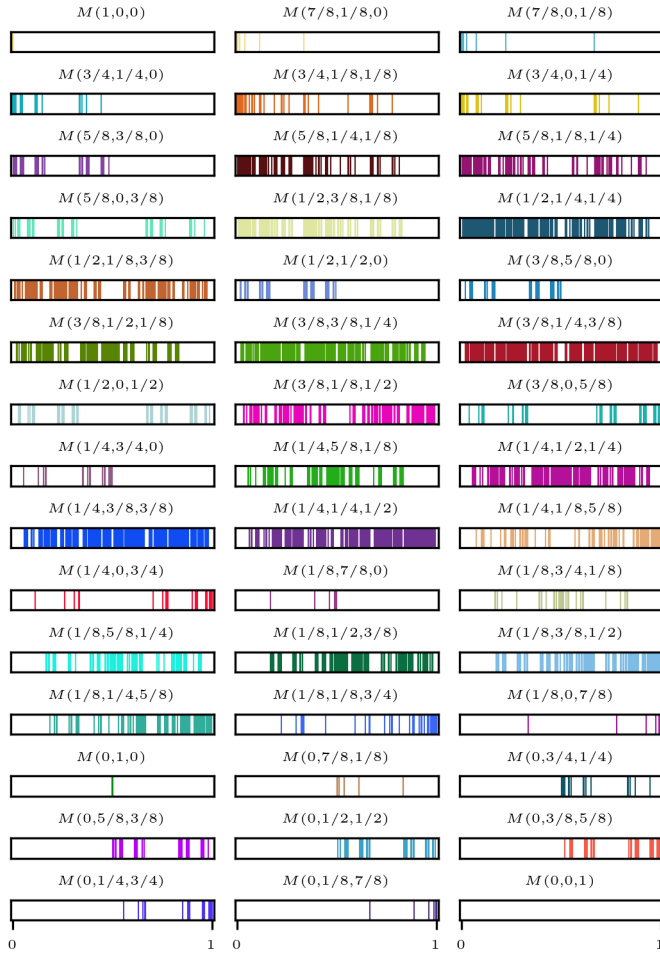
$$\alpha(\omega_n) = \frac{\log \mu[C_{\omega_n}]}{\log \delta_n} = - \sum_{i \in \Omega} \Phi_i(\omega_n) \log_b p_i.$$

En el límite  $n \rightarrow \infty$ ,  $\alpha(\omega_n)$  tiende a un valor bien definido para casi todo punto, caracterizando la singularidad local de la medida.

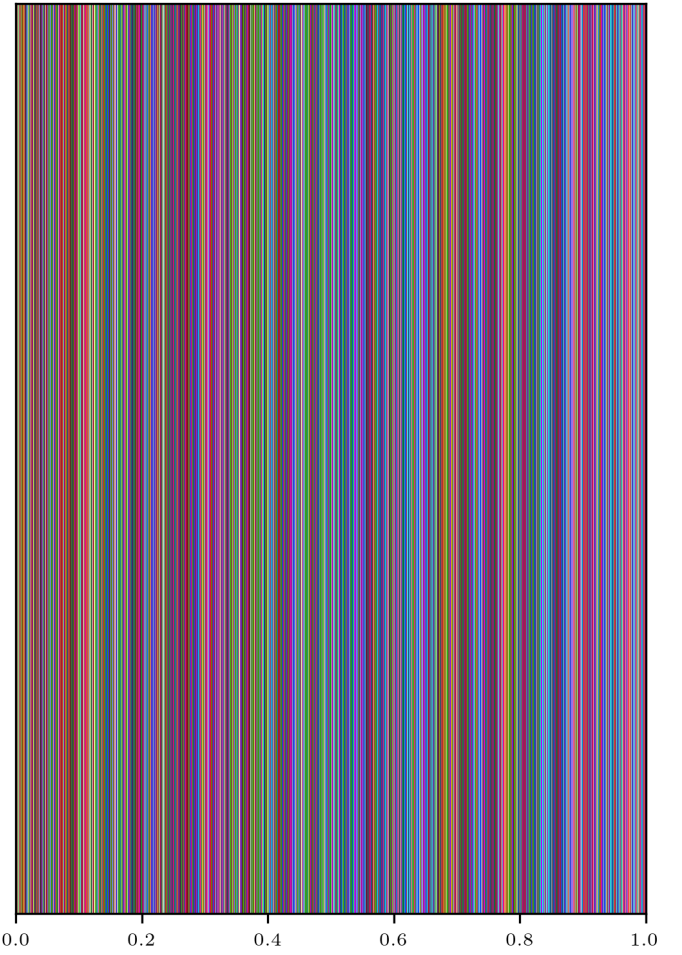
Un aspecto sutil es que la relación entre el vector de frecuencias  $\phi$  y el exponente  $\alpha$  no es inyectiva. Fijado  $\alpha$ , la ecuación

$$\alpha = - \sum_{i \in \Omega} \phi_i \log_b p_i, \quad \text{con} \quad \sum_i \phi_i = 1,$$

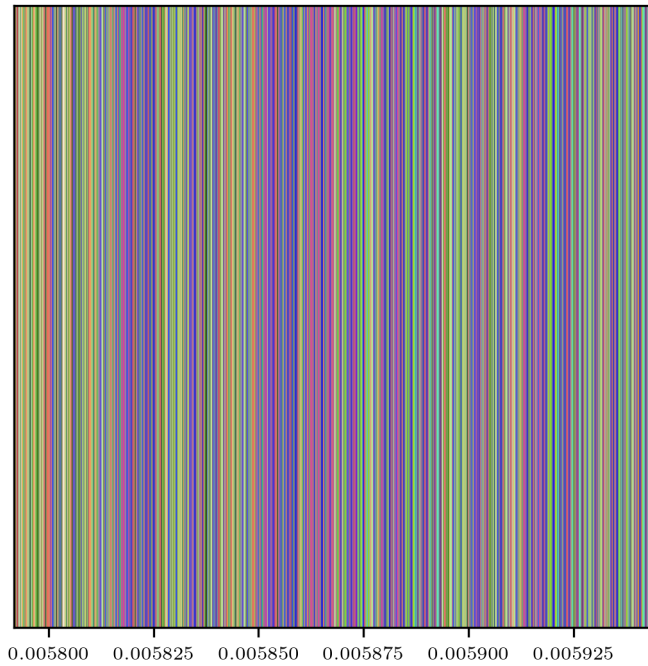
define un hiperplano (o una recta, para  $b = 3$ ) en el espacio de frecuencias. Por lo tanto, hay infinitos conjuntos  $M(\phi)$  distintos que comparten el mismo exponente de Hölder  $\alpha$ .



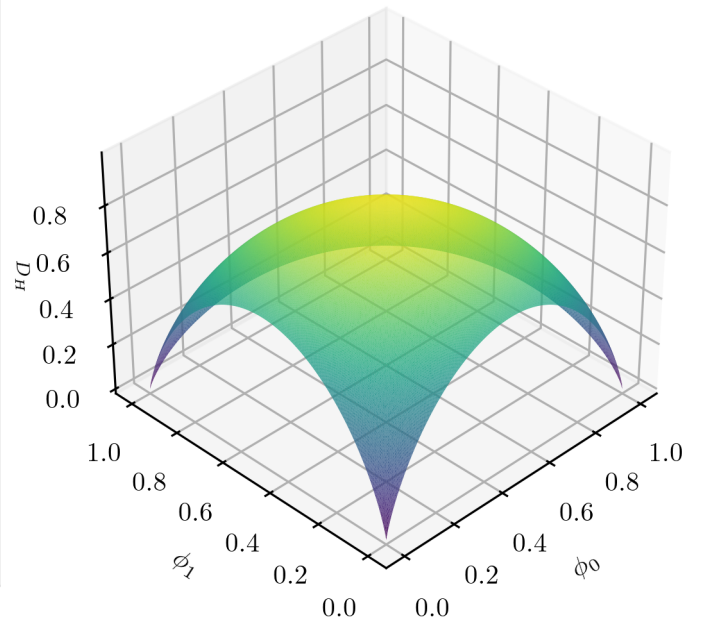
**Figura 3.13:** Los 45 conjuntos  $M(\phi_0, \phi_1, \phi_2)$  para  $n = 8$ , cada uno coloreado según su dimensión fractal.



**Figura 3.14:** Espectro óptico resultante de superponer todos los conjuntos de la figura anterior.



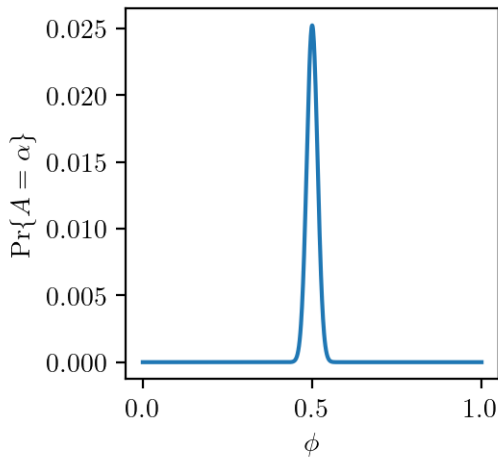
**Figura 3.15:** Espectro óptico dentro de un solo cilindro de orden 8, mostrando autosimilitud.



**Figura 3.16:** Superficie de dimensión  $D_H(\phi_0, \phi_1, \phi_2)$  para  $b = 3$ .

El conjunto  $J_\alpha$  de todos los puntos con exponente  $\alpha$  es entonces una unión (posiblemente no numerable) de conjuntos  $M(\phi)$  que satisfacen la ecuación anterior. Esta es la esencia de la *separación multifractal*: el soporte de la medida se descompone en una jerarquía de subconjuntos fractales, cada uno caracterizado por un  $\alpha$ , pero cada uno de esos subconjuntos puede a su vez ser una mezcla de diferentes “tipo” estadísticos  $\phi$ .

La distribución de probabilidad del exponente  $\alpha$  para  $n$  grande se concentra en torno al valor  $\alpha_0 = -\sum_i (1/s) \log_b p_i$  (cuando todas las probabilidades activas son iguales), como muestra la Figura 3.17 para el caso binario.



**Figura 3.17:** Distribución de probabilidad del exponente  $\alpha$  para una medida binomial con  $n = 1000$ . El pico corresponde al valor más probable, asociado a frecuencias equilibradas.

Esta rica estructura local y global es lo que hace de la medida multinomial un objeto de estudio central en la teoría multifractal, sirviendo como puente entre la probabilidad, la teoría de la información y la geometría fractal.

#### 3.5.4. Vínculos con la Mecánica Estadística de Boltzmann

El conjunto  $J_\alpha$ , definido como la unión de todos los conjuntos  $M(\phi_0, \phi_1, \dots, \phi_{b-1})$  que comparten el mismo exponente de Hölder  $\alpha$ , posee inherentemente una estructura multifractal. Cada componente  $M(\cdot)$  es en sí mismo un fractal, cuya dimensión está dada por el teorema de Eggleston. Para determinar la dimensión de Hausdorff global de  $J_\alpha$ , se recurre a dos propiedades fundamentales de esta dimensión:

$$D_H(M) \leq D_H(M') \quad \text{si } M \subset M', \quad (3.19)$$

$$D_H\left(\bigcup_n M_n\right) = \sup_n D_H(M_n). \quad (3.20)$$

En consecuencia, la dimensión de Hausdorff de  $J_\alpha$  será el supremo (o máximo, si se alcanza) de las dimensiones de sus subconjuntos constituyentes.

Denotando por  $D(\alpha)$  esta dimensión, se tiene:

$$D(\alpha) = D_H(J_\alpha) = \max_{(\phi')} D_H(M(\phi'_0, \phi'_1, \dots, \phi'_{b-1})) = \frac{1}{\log b} \max_{(\phi')} S(\phi'_0, \phi'_1, \dots, \phi'_{b-1}), \quad (3.21)$$

donde  $S(\phi) = -\sum_r \phi_r \log \phi_r$  es la entropía de Shannon (hasta una constante multiplicativa).

Las propiedades (3.19) y (3.20), junto con el teorema de Eggleston, permiten establecer un paralelismo profundo entre la teoría multifractal y las ideas fundamentales de la mecánica estadística de



Boltzmann, particularmente en lo concerniente a la irreversibilidad y la búsqueda del estado de máxima entropía. Este paralelismo provee un marco poderoso para determinar el espectro de dimensiones de un multifractal de naturaleza estadística.

Recordemos brevemente el razonamiento de Boltzmann para un sistema aislado de  $N$  partículas no interactuantes. El estado microscópico se representa como una nube de  $N$  puntos en el espacio  $\mu$  (espacio de fases de una partícula). Al dividir este espacio en  $K$  celdas, un estado mesoscópico se describe mediante el vector de ocupación  $\{n\} = (n_1, n_2, \dots, n_K)$ . La entropía asociada a este estado es proporcional al logaritmo del número de microestados compatibles:

$$S(\{n\}) = k \log \frac{N!}{n_1! n_2! \dots n_K!}.$$

Para  $N$  grande, usando la aproximación de Stirling y definiendo las fracciones  $\phi_j = n_j/N$ , esta entropía se expresa como  $S(\{n\}) \approx kN \sum_j \phi_j \log \phi_j$ , que no es más que la entropía de Shannon (salvo constantes). En el límite continuo ( $K \rightarrow \infty$ ),  $\phi_j$  se convierte en una función de distribución  $f(\mathbf{r}, \mathbf{v}, t)$  gobernada por la ecuación de Boltzmann. El célebre teorema  $H$  establece que esta entropía aumenta monótonamente en el tiempo, alcanzando un máximo en el equilibrio, que corresponde a la distribución de Maxwell-Boltzmann.

Inspirados por esta construcción, para hallar  $D(\alpha)$  maximizamos la entropía de Shannon  $S(\phi') = -\sum_{r=0}^{b-1} \phi'_r \log \phi'_r$ , pero ahora sujeto a dos restricciones lineales que definen el conjunto  $J_\alpha$ :

$$\alpha = -\sum_{i \in \Omega} \phi_i \log_b p_i, \quad (3.22)$$

$$1 = \sum_{i \in \Omega} \phi_i. \quad (3.23)$$

Nótese la analogía termodinámica: la primera restricción es análoga a fijar la energía media por “estado”  $r$ , si identificamos el “nivel de energía” como  $\varepsilon_r = -\log_b p_r$ . La segunda es la condición de normalización.

Resolviendo este problema de maximización con restricciones mediante multiplicadores de Lagrange, se encuentra que la distribución de frecuencias óptima es la distribución de Gibbs:

$$\phi'_r = P_r(q) = \frac{p_r^q}{Z_q}, \quad (3.24)$$

donde  $Z_q = \sum_{r=0}^{b-1} p_r^q$  actúa como una función de partición, y  $q$  es un multiplicador de Lagrange determinado implícitamente por la restricción del valor de  $\alpha$ :

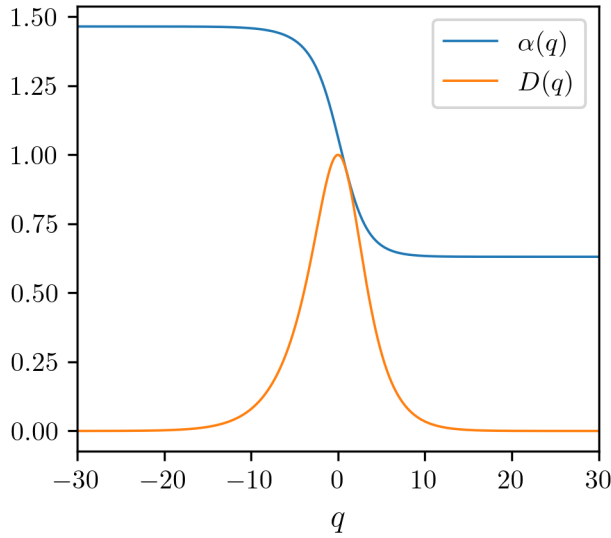
$$\alpha(q) = \sum_{r=0}^{b-1} P_r(q) \left( -\frac{\log p_r}{\log b} \right) \equiv \left\langle -\frac{\log p}{\log b} \right\rangle_q. \quad (3.25)$$

El símbolo  $\langle \cdot \rangle_q$  denota el promedio ponderado con la distribución  $P_r(q)$ .

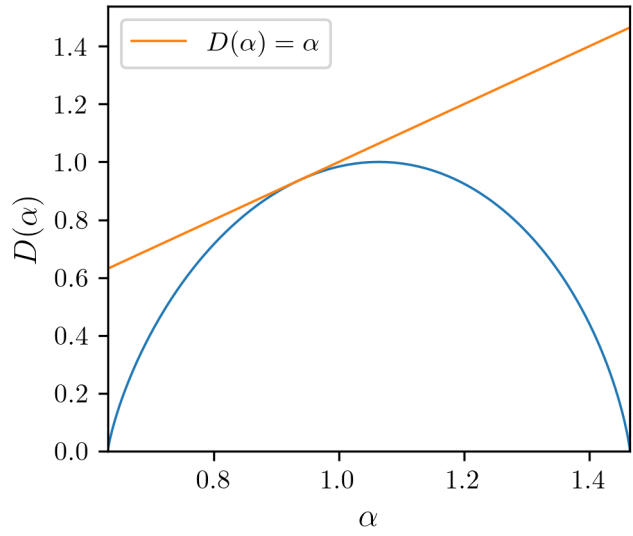
Sustituyendo la distribución óptima (3.24) en la expresión de la entropía, obtenemos la dimensión de Hausdorff de  $J_{\alpha(q)}$ :

$$D(q) = D(\alpha(q)) = -\frac{1}{\log b} \sum_{r=0}^{b-1} P_r(q) \log P_r(q) = \left\langle -\frac{\log P_q}{\log b} \right\rangle_q. \quad (3.26)$$

La Figura 3.18 ilustra el comportamiento de  $D(q)$  y  $\alpha(q)$  para un caso concreto ( $p_0 = 0.2, p_1 = 0.3, p_2 = 0.5$ ). Eliminando el parámetro  $q$  de las ecuaciones (3.25) y (3.26), se obtiene la curva  $D(\alpha)$ , que es el espectro de dimensiones multifractal, mostrado en la Figura 3.19.



**Figura 3.18:** Variación de  $D(q)$  y  $\alpha(q)$  con el parámetro  $q$ , para el vector de probabilidades  $(0.2, 0.3, 0.5)$ .



**Figura 3.19:** Espectro de dimensiones  $D(\alpha)$  resultante para el mismo conjunto de probabilidades.

Es posible relacionar  $\alpha(q)$  directamente con la función de partición  $Z_q$ . Observando que

$$\frac{d}{dq} Z_q = \frac{d}{dq} \sum_r p_r^q = \sum_r p_r^q \log p_r,$$

y usando la definición (3.25), se llega a:

$$\alpha(q) = -\frac{1}{\log b} \frac{d}{dq} \log Z_q = \frac{d}{dq} \left( -\frac{\log Z_q}{\log b} \right) \equiv \frac{d\tau(q)}{dq}, \quad (3.27)$$

donde hemos definido la función auxiliar

$$\tau(q) = -\frac{\log Z_q}{\log b}. \quad (3.28)$$

La relación entre  $D(\alpha)$  y  $\alpha(q)$  se obtiene sustituyendo  $P_r(q)$  en (3.26):

$$D(\alpha(q)) = -\frac{1}{\log b} \sum_r P_r(q) (q \log p_r - \log Z_q) = q\alpha(q) - \tau(q). \quad (3.29)$$

Derivando esta última expresión respecto a  $\alpha$  y empleando (3.27), se encuentra una relación simple:

$$\frac{dD(\alpha)}{d\alpha} = q. \quad (3.30)$$

Las ecuaciones (3.27), (3.29) y (3.30) revelan que  $D(\alpha)$  y  $\tau(q)$  están relacionadas mediante una *transformada de Legendre*. En efecto, de (3.30) tenemos  $dD = q d\alpha = d(q\alpha) - \alpha dq$ , por lo que

$$d(D - q\alpha) = -\alpha dq.$$

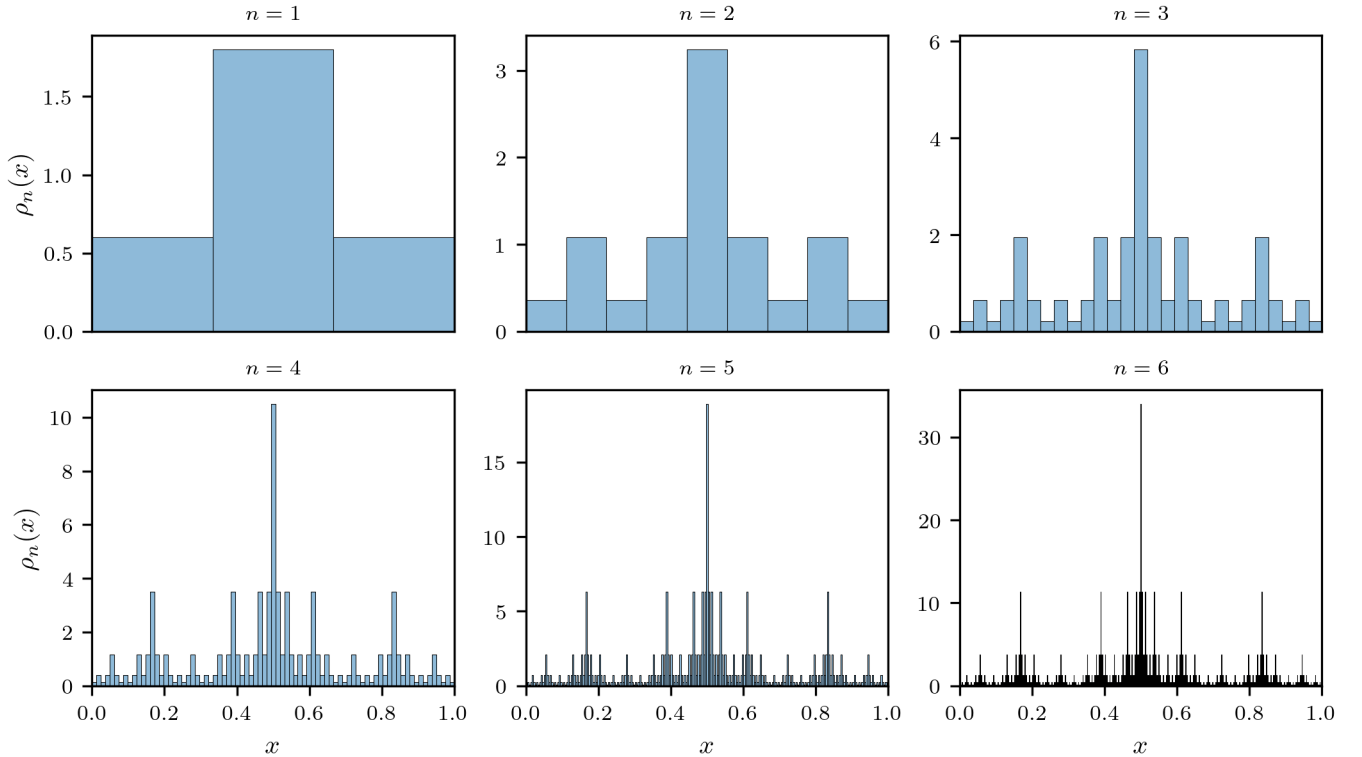
Comparando con la diferencial de (3.29),  $d(D - q\alpha) = -d\tau$ , concluimos que  $d\tau = \alpha dq$ , que es precisamente (3.27). Por lo tanto,  $D(\alpha)$  y  $\tau(q)$  contienen la misma información geométrica, siendo una la transformada de Legendre de la otra:  $\tau(q) = \inf_{\alpha}[q\alpha - D(\alpha)]$  y  $D(\alpha) = \inf_q[q\alpha - \tau(q)]$ .

### Ejemplo Ilustrativo [44]

Consideremos un IFS definido sobre el intervalo unitario  $I = [0, 1]$  con tres contracciones lineales de factor  $1/3$  y probabilidades  $p_0 = p_2$  y  $p_1 = 1 - 2p_0$ , con la condición  $p_0 < p_1$  (la probabilidad se concentra en el intervalo central). Formalmente:

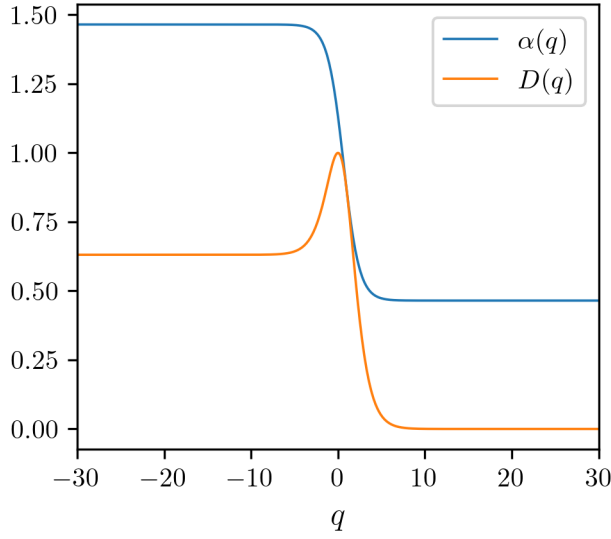
$$\left\{ X = [0, 1]; w_0(x) = \frac{x}{3}, w_1(x) = \frac{x}{3} + \frac{1}{3}, w_2(x) = \frac{x}{3} + \frac{2}{3}; p_0, p_1, p_2 = p_0 \right\}.$$

La Figura 3.20 muestra las primeras iteraciones de la construcción de esta medida.

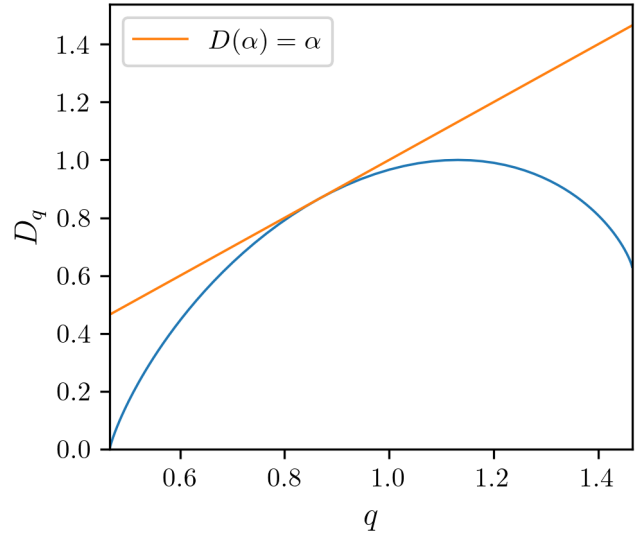


**Figura 3.20:** Primeras seis iteraciones en la construcción de una medida fractal sobre  $[0, 1]$  con mayor probabilidad en el tercio central.

Aplicando las fórmulas generales (3.25) y (3.26) a este caso particular (por ejemplo, con  $p_0 = 0.2, p_1 = 0.6$ ), se obtienen las curvas de  $\alpha(q)$  y  $D(q)$  mostradas en la Figura 3.21. El espectro de dimensiones  $D(\alpha)$ , que resulta de eliminar el parámetro  $q$ , se presenta en la Figura 3.22.



**Figura 3.21:** Comportamiento de  $D(q)$  y  $\alpha(q)$  para  $p_0 = p_2 = 0.2$  y  $p_1 = 0.6$ .



**Figura 3.22:** Espectro multifractal  $D(\alpha)$  correspondiente a los parámetros anteriores.

### 3.5.5. El Conjunto de Condensación de una Medida

Retomando la definición de  $J_\alpha$ , es de interés calcular su *peso estadístico* o medida total. Recordemos que un cilindro de orden  $K$ , denotado  $C_{z_1 z_2 \dots z_K}$ , es un intervalo de longitud  $\Delta = b^{-K}$  asociado a una secuencia finita de dígitos. Su medida, usando el vector de probabilidades original  $\mathbf{p} = (p_i)$ , es:

$$\mu(C_{z_1 z_2 \dots z_K}) = p_0^{n_0(K)} p_1^{n_1(K)} \dots p_{b-1}^{n_{b-1}(K)} = (\Delta)^\alpha,$$

donde  $n_i(K)$  es el número de apariciones del dígito  $i$  en la secuencia, y  $\alpha$  es el exponente de Hölder local para ese cilindro.

Si consideramos específicamente cilindros cuyos puntos tienen exponente de Hölder igual a un valor fijo  $\alpha_0$ , su medida individual será  $\mu_{\alpha_0}(C) = \Delta^{\alpha_0}$ . Sea  $N_{\alpha_0}(\Delta)$  el número de tales cilindros de tamaño  $\Delta$  necesarios para cubrir  $J_{\alpha_0}$ . Por definición de dimensión de Hausdorff, para  $\Delta$  pequeño se tiene:

$$N_{\alpha_0}(\Delta) \sim \Delta^{-D(\alpha_0)},$$

donde  $D(\alpha_0) = D_H(J_{\alpha_0})$ . Por lo tanto, la medida total de la cubierta de  $J_{\alpha_0}$  con cilindros de tamaño  $\Delta$  escala como:

$$\mu_{J_{\alpha_0}}(\Delta) = N_{\alpha_0}(\Delta) \cdot \mu_{\alpha_0}(C) \sim \Delta^{\alpha_0 - D(\alpha_0)}.$$

Este resultado es crucial: cuando  $\Delta \rightarrow 0$ , el comportamiento asintótico de  $\mu_{J_{\alpha_0}}(\Delta)$  depende del signo del exponente  $\alpha_0 - D(\alpha_0)$ . Si  $\alpha_0 > D(\alpha_0)$ , la medida total tiende a cero; si  $\alpha_0 < D(\alpha_0)$ , diverge (pero está acotada por 1, siendo una probabilidad). La medida total de  $J_\alpha$  (en el límite de resolución infinita) será no nula únicamente para aquel valor especial  $\alpha_c$  que satisface la condición de balance:

$$\alpha_c = D(\alpha_c).$$

Para cualquier otro  $\alpha$ ,  $\mu(J_\alpha) = 0$  en el límite. En otras palabras, la medida original  $\mu$  se *condensa* en el conjunto  $J_{\alpha_c}$ :

$$\mu(J_\alpha) = \delta(\alpha - \alpha_c),$$

donde  $\delta$  es la delta de Dirac. Al conjunto  $J_{\alpha_c}$  se le denomina *conjunto de condensación* de la medida  $\mu$ .

### 3.5.6. Reescalado con una Media Distorsionada: $P(q)$

Surge una pregunta natural: de todos los conjuntos  $J_\alpha$  definidos por el vector original  $\mathbf{p}$ , ¿cuál adquiere un peso significativo si, en lugar de  $\mathbf{p}$ , utilizamos para el cálculo de la medida un *diferente* vector de probabilidades  $P(q)$ ? Este nuevo vector se define precisamente como la distribución de Gibbs que maximiza la entropía:

$$P(q) = (P_0(q), P_1(q), \dots, P_{b-1}(q)), \quad \text{con } P_r(q) = \frac{p_r^q}{Z_q}.$$

La medida de un cilindro, calculada ahora con  $P(q)$ , es:

$$\mu(q, C_{z_1 z_2 \dots z_K}) = P_0(q)^{n_0(K)} P_1(q)^{n_1(K)} \dots = \frac{[\mu(C)]^q}{Z_q^K}.$$

Dado que  $Z_q^K = b^{K \log_b Z_q} = \Delta^{-\tau(q)}$  con  $\tau(q) = -\log_b Z_q$ , podemos relacionar las dos medidas:

$$\mu(q, C) = \Delta^{\tau(q)} [\mu(C)]^q.$$

Para un cilindro típico de un punto con exponente  $\alpha_0$ ,  $\mu(C) = \Delta^{\alpha_0}$ , luego:

$$\mu_{\alpha_0}(q, C) = \Delta^{\tau(q) + q\alpha_0}.$$

Siguiendo un razonamiento similar al de la sección anterior, la medida total (con  $P(q)$ ) de una cubierta de  $J_{\alpha_0}$  escala como:

$$\mu_{J_{\alpha_0}}(q, \Delta) \sim \Delta^{q\alpha_0 - \tau(q) + D(\alpha_0)}.$$

Para que esta medida total no sea ni cero ni infinita en el límite  $\Delta \rightarrow 0$  (es decir, para que el conjunto  $J_\alpha$  porte una fracción finita de la medida reescalada), el exponente debe anularse. Esto conduce a la condición:

$$q\alpha - D(\alpha) = \tau(q). \quad (3.31)$$

Dado que  $\tau(q)$  se define como el ínfimo de  $q\alpha - D(\alpha)$  (transformada de Legendre), la igualdad (3.31) se satisface precisamente para el valor  $\alpha^*(q)$  que realiza el mínimo (o, equivalentemente, donde la derivada es cero). Así, para un  $q$  fijo, cuando la resolución se hace infinitamente fina ( $\Delta \rightarrow 0$ ), la medida reescalada  $\mu(q, \cdot)$  se concentra completamente en el conjunto  $J_{\alpha^*(q)}$ .

Podemos analizar con más detalle cómo ocurre esta concentración. La medida  $\mu(q, J_\alpha(\Delta))$  como función de  $\alpha$  para  $\Delta$  fijo tiene un máximo en un valor  $\alpha^*$  que satisface:

$$\left. \frac{\partial}{\partial \alpha} \log \mu(q, J_\alpha(\Delta)) \right|_{\alpha^*} = \left( q - \left. \frac{dD}{d\alpha} \right|_{\alpha^*} \right) \log \Delta = 0,$$

lo cual implica  $q = dD/d\alpha|_{\alpha^*}$ , reconociendo a  $q$  como la pendiente de la curva  $D(\alpha)$  en ese punto. La segunda derivada, suponiendo  $D(\alpha)$  cóncava ( $D''(\alpha) < 0$ ), es negativa, confirmando que es un máximo.

Expandimos alrededor de este máximo para obtener la forma de la distribución:

$$\mu(q, J_\alpha(\Delta)) \propto \exp \left( \frac{(\alpha - \alpha^*)^2}{2\sigma_\alpha^2} \right), \quad \text{con } \sigma_\alpha^2 = \frac{1}{|D''(\alpha^*)| |\log \Delta|}.$$

La varianza  $\sigma_\alpha^2$  disminuye a medida que  $\Delta \rightarrow 0$  (pues  $|\log \Delta| \rightarrow \infty$ ), lo que significa que el pico de la distribución se hace cada vez más agudo. Si consideramos la medida acumulada en una banda  $S_\epsilon(\Delta) = \bigcup_{|\alpha - \alpha^*| < \epsilon} J_\alpha(\Delta)$ , un cálculo integral muestra que, en el límite  $\Delta \rightarrow 0$ , esta medida tiende a 1 para cualquier  $\epsilon > 0$  fijo. En otras palabras, toda la medida  $\mu(q, \cdot)$  se condensa en el conjunto correspondiente a  $\alpha^*$ .

Este es el *teorema de condensación de la medida*: para cada valor del parámetro  $q$ , existe un exponente de Hölder característico  $\alpha^*(q)$  tal que

$$\lim_{\Delta \rightarrow 0} \mu(q, J_\alpha(\Delta)) = \delta(\alpha - \alpha^*(q)).$$

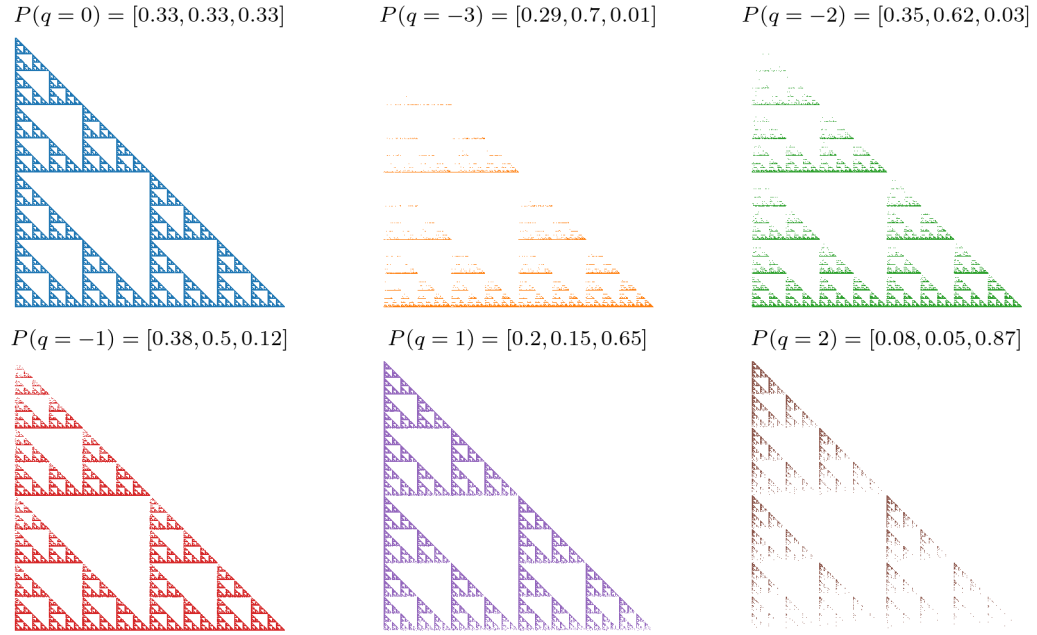
El conjunto  $J_{\alpha^*(q)}$  es el conjunto de condensación para la medida distorsionada por el parámetro  $q$ .

En la práctica, para un  $q$  dado, el procedimiento es el siguiente: 1. Calcular la función de partición  $Z_q = \sum_r p_r^q$  y el vector  $P_r(q) = p_r^q / Z_q$ . 2. Obtener  $\tau(q) = -\log_b Z_q$ . 3. El exponente de condensación es  $\alpha^*(q) = \sum_r P_r(q) (-\log_b p_r)$ . 4. La dimensión del conjunto de condensación es  $D(\alpha^*(q)) = q\alpha^*(q) - \tau(q)$ .

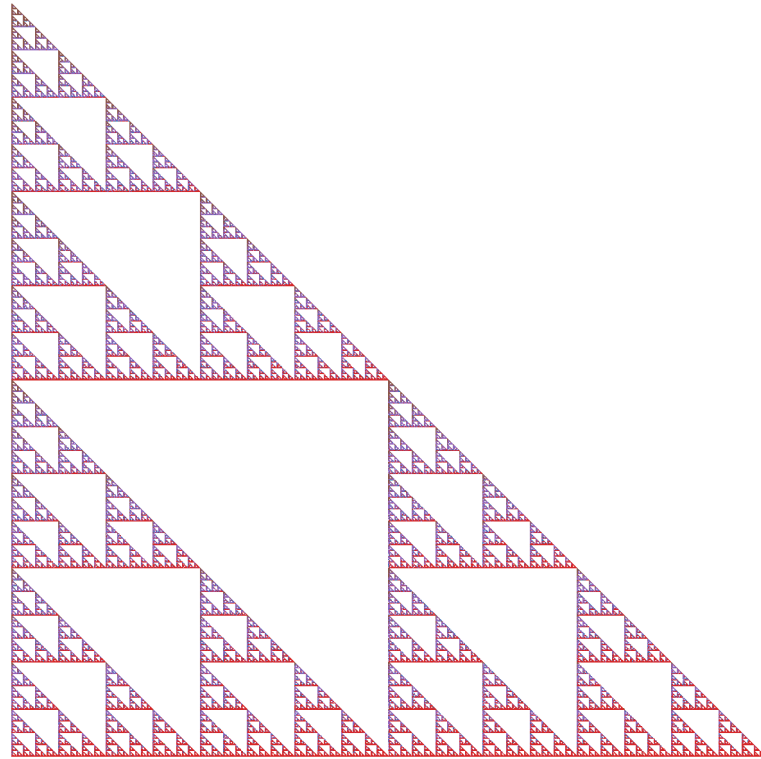
Este último resultado puede interpretarse también como el cálculo de la transformada de Legendre de  $\tau(q)$ . Visualmente, el conjunto de condensación para un  $q$  dado puede generarse utilizando el IFS original pero reemplazando las probabilidades originales  $\mathbf{p}$  por las probabilidades reescaladas  $P(q)$ . Las Figuras 3.23 y 3.24 muestran ejemplos de tales conjuntos para diferentes valores de  $q$ , en el contexto de un triángulo de Sierpinski.

## 3.6. Caracterización mediante Momentos Globales y Dualidad de Legendre

Una perspectiva complementaria para analizar la medida  $\mu$  consiste en examinar sus momentos globales de orden  $q$ . Este enfoque, basado en promedios estadísticos, resulta ser dual a la descripción



**Figura 3.23:** Representación del conjunto de condensación (análogo a un triángulo de Sierpinski ponderado) para distintos valores del parámetro  $q$ , que modifica las probabilidades de las contracciones.



**Figura 3.24:** Superposición de los conjuntos mostrados en la figura anterior, ilustrando cómo la medida se condensa en diferentes subconjuntos fractales al variar  $q$ .

local mediante el exponente de Hölder  $\alpha$ . A continuación, se establece esta dualidad a través de la transformada de Legendre.

### 3.6.1. El Exponente de Correlación $\tau(q)$ como Momento Escalante

En el contexto de la medida multinomial, el exponente de correlación introducido en (3.9) puede reformularse explícitamente como:

$$\tau(q) = \lim_{n \rightarrow \infty} \frac{\log \left( \sum_{\omega \in \Omega_0^n} \mu^q[C_{\omega_n}(\omega)] \right)}{\log \delta_n}, \quad (3.32)$$

con  $q \in \mathbb{R}$  y  $\delta_n = b^{-n}$  denotando la longitud de los cilindros de orden  $n$ . Desarrollando la suma sobre todas las secuencias posibles, obtenemos:

$$\begin{aligned} \tau(q) &= \lim_{n \rightarrow \infty} \frac{-1}{n} \log_b \left[ \sum_{w_1 \in \Omega} \cdots \sum_{w_n \in \Omega} p_{w_1}^q \cdots p_{w_n}^q \right] \\ &= -\log_b \left( \sum_{\omega \in \Omega} p_{\omega}^q \right). \end{aligned} \quad (3.33)$$

La cantidad  $Z_q = \sum_{\omega \in \Omega} p_{\omega}^q$  juega un papel análogo a la *función de partición* en mecánica estadística, agrupando la información de los pesos probabilísticos elevados a la potencia  $q$ .

#### Ejemplo Ilustrativo con Base $b = 10$

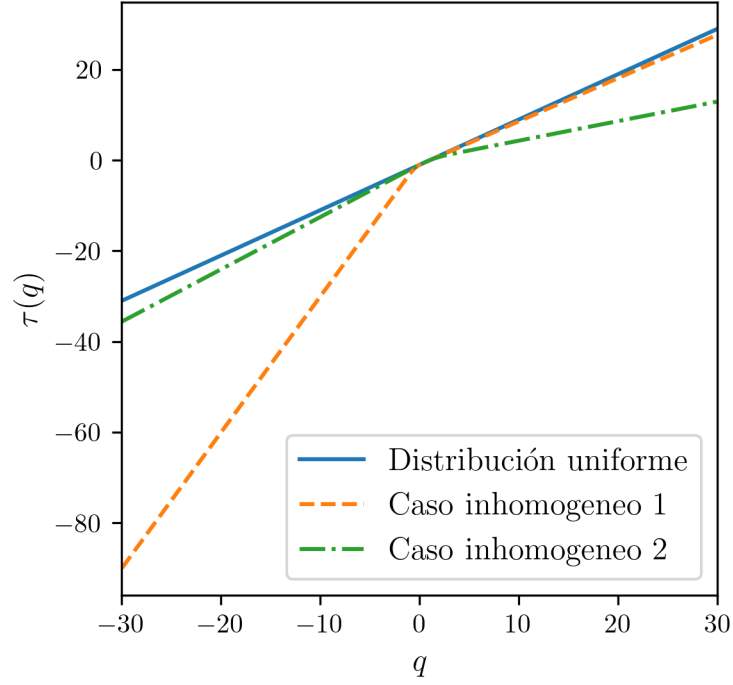
Consideremos una medida multinomial con base  $b = 10$ . Para ilustrar cómo  $\tau(q)$  refleja la inhomogeneidad de la medida, examinemos tres escenarios distintos:

1. Distribución uniforme:  $p_{\omega} = 0.1$  para todo  $\omega = 0, \dots, 9$ . En este caso,  $\tau(q)$  es una función lineal:  $\tau(q) = (q - 1) \log_b 10 = (q - 1)$ . Esta recta aparece como la línea sólida en la Figura 3.25.
2. Presencia de un peso muy pequeño: Todos los  $p_{\omega}$  valen 0.111 excepto uno que es 0.001. Esta configuración introduce regiones con medida extremadamente baja. La curva  $\tau(q)$  resultante (línea discontinua) se desvía de la lineal para  $q < 0$ , mostrando una pendiente más pronunciada, mientras que para  $q > 0$  se mantiene cercana al caso uniforme.
3. Presencia de un peso dominante: Todos los  $p_{\omega}$  valen 0.07 excepto uno que es 0.37. Aquí existe una región con medida relativamente alta. La curva  $\tau(q)$  (línea punteada) se separa de la lineal para  $q > 0$ , cayendo más rápidamente, mientras que para  $q < 0$  se aproxima al comportamiento uniforme.

En general, cuando la medida es uniforme sobre su soporte (todos los  $p_{\omega}$  iguales y no nulos),  $\tau(q)$  es una línea recta de pendiente  $D_0$ , la dimensión del soporte. En cambio, una medida no uniforme produce una  $\tau(q)$  cóncava, siempre creciente, cuyos extremos asintóticos están anclados en los valores  $q = 0$  y  $q \rightarrow \pm\infty$ . El teorema 3.1.1 cuantifica precisamente estas desigualdades.

Interpretativamente, para  $q > 0$  el comportamiento de  $\tau(q)$  está dominado por las regiones donde la medida es relativamente grande, mientras que para  $q < 0$  está controlado por las regiones donde





**Figura 3.25:** Función  $\tau(q)$  para una medida multinomial con  $b = 10$ . La línea sólida corresponde al caso uniforme ( $p_\omega = 0.1$ ). La línea discontinua representa un caso con un peso muy pequeño (0.001), y la punteada un caso con un peso dominante (0.37).

la medida es positiva pero muy pequeña. Así,  $\tau(q)$  cuantifica globalmente las desviaciones de la homogeneidad.

### 3.6.2. La Transformada de Legendre como Puente entre lo Local y lo Global

El objetivo central ahora es demostrar que la relación entre el espectro multifractal  $f(\alpha)$  (descripción local) y el exponente de correlación  $\tau(q)$  (descripción global) constituye una *transformada de Legendre*. Partimos de la expresión de  $f(\alpha)$  como la dimensión de Hausdorff del conjunto de puntos con exponente  $\alpha$ , dada por el teorema de Eggleston:

$$f(\alpha) = -\frac{1}{\log b} \sum_{\omega \in \Omega} \phi_\omega \log_b \phi_\omega,$$

donde  $\phi_\omega$  son las frecuencias asintóticas de los dígitos.

El valor  $f(\alpha)$  corresponde al máximo de esta expresión sujeto a dos restricciones lineales que definen el conjunto  $J_\alpha$ :

$$\sum_{\omega \in \Omega} \phi_\omega = 1, \tag{3.34}$$

$$\alpha = -\sum_{\omega \in \Omega} \phi_\omega \log_b p_\omega. \tag{3.35}$$

### 3.6. CARACTERIZACIÓN MEDIANTE MOMENTOS GLOBALES Y DUALIDAD DE LEGENDRE

Para realizar esta optimización, introducimos dos multiplicadores de Lagrange,  $\lambda$  y  $q$ , y definimos el funcional:

$$Q = - \sum_{\omega \in \Omega} \phi_{\omega} \log_b \phi_{\omega} + q \left( \alpha + \sum_{\omega \in \Omega} \phi_{\omega} \log_b p_{\omega} \right) + \lambda \left( -1 + \sum_{\omega \in \Omega} \phi_{\omega} \right).$$

Derivando respecto a  $\phi_{\omega}$  e igualando a cero:

$$\frac{\partial Q}{\partial \phi_{\omega}} = -\log_b \phi_{\omega} - 1 + q \log_b p_{\omega} + \lambda = 0, \quad \omega \in \Omega. \quad (3.36)$$

De aquí despejamos  $\phi_{\omega}$ :

$$\log_b \phi_{\omega} = q \log_b p_{\omega} + (\lambda - 1) \Rightarrow \phi_{\omega} = b^{\lambda-1} p_{\omega}^q.$$

Imponiendo la restricción de normalización (3.34):

$$\sum_{\omega \in \Omega} b^{\lambda-1} p_{\omega}^q = 1 \Rightarrow b^{\lambda-1} = \frac{1}{\sum_{i \in \Omega} p_i^q}.$$

Por lo tanto, las frecuencias óptimas (que maximizan la entropía) son:

$$\phi_{\omega} = \frac{p_{\omega}^q}{\sum_{i \in \Omega} p_i^q}, \quad \omega \in \Omega. \quad (3.37)$$

Sustituyendo (3.37) en la restricción (3.35), obtenemos el valor de  $\alpha$  correspondiente a este óptimo, denotado  $\alpha_q$ :

$$\alpha_q = -\frac{1}{\sum_{i \in \Omega} p_i^q} \sum_{\omega \in \Omega} p_{\omega}^q \log_b p_{\omega}. \quad (3.38)$$

Finalmente, el valor máximo de la entropía (que es  $f(\alpha_q)$ ) se calcula sustituyendo (3.37) en la expresión de Eggleston:

$$\begin{aligned} f(\alpha_q) &= - \sum_{\omega \in \Omega} \frac{p_{\omega}^q}{\sum_i p_i^q} \log_b \left( \frac{p_{\omega}^q}{\sum_j p_j^q} \right) \\ &= - \frac{q}{\sum_i p_i^q} \sum_{\omega \in \Omega} p_{\omega}^q \log_b p_{\omega} + \log_b \left( \sum_j p_j^q \right) \\ &= q\alpha_q + \tau(q), \end{aligned} \quad (3.39)$$

donde hemos usado la definición  $\tau(q) = -\log_b \sum_j p_j^q$ . Reordenando, obtenemos la relación fundamental:

$$\tau(q) = q\alpha_q - f(\alpha_q). \quad (3.40)$$

Esta es precisamente la forma de una *transformada de Legendre*:  $\tau(q) = \inf_{\alpha} \{q\alpha - f(\alpha)\}$ , donde el ínfimo se alcanza en  $\alpha = \alpha_q$ . La función  $\alpha_q$  proporciona, para cada  $q$ , el exponente de Hölder para el cual el conjunto correspondiente posee la máxima dimensión fractal. Los límites asintóticos:

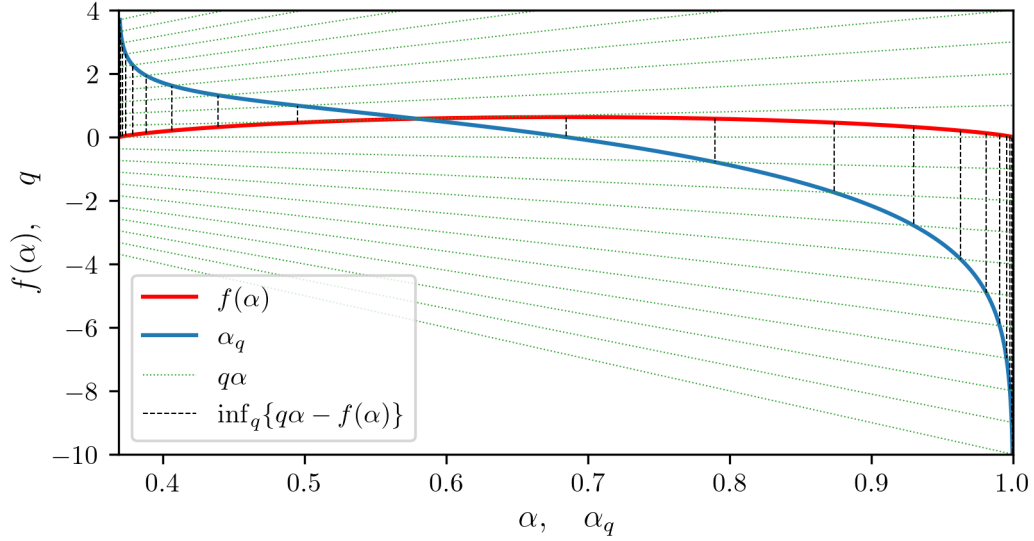
$$\alpha_{\min} = \lim_{q \rightarrow +\infty} \alpha_q = \min_{\omega \in \Omega} (-\log_b p_{\omega}), \quad (3.41)$$

$$\alpha_{\max} = \lim_{q \rightarrow -\infty} \alpha_q = \max_{\omega \in \Omega} (-\log_b p_{\omega}), \quad (3.42)$$

corresponden a los exponentes más extremos presentes en la medida.

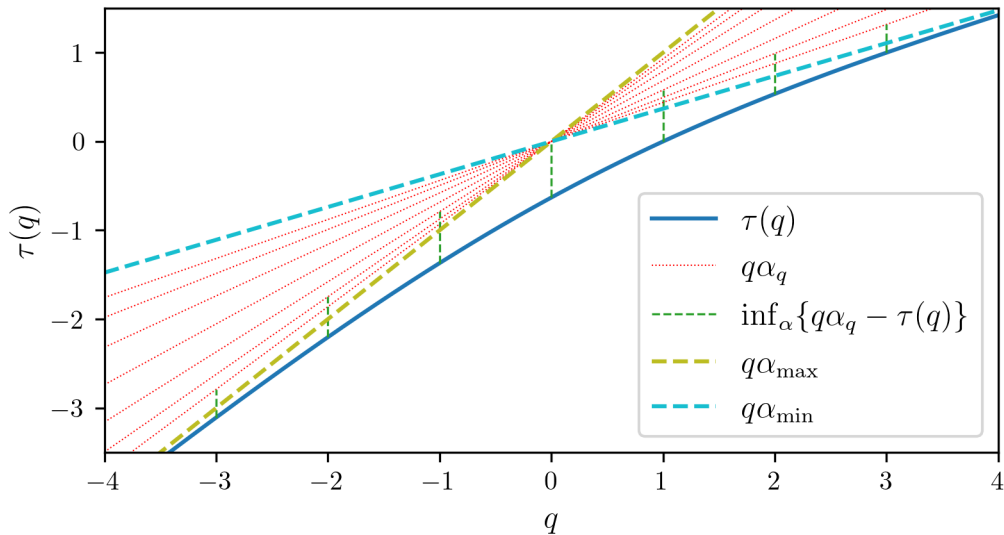
### Ejemplo: Medida de Cantor

Consideremos el caso clásico con  $b = 3$ ,  $p_0 = 1/3$ ,  $p_1 = 0$  y  $p_2 = 2/3$ . La Figura 3.26 muestra la curva  $f(\alpha)$  y las rectas  $q\alpha$  para distintos  $q$ . El valor  $\alpha_q$  en cada caso es la abscisa donde la distancia vertical entre la recta y la curva es mínima (es decir, donde la recta tangente a  $f(\alpha)$  tiene pendiente  $q$ ).



**Figura 3.26:** Transformada de Legendre para la medida de Cantor. La curva sólida es  $f(\alpha)$ . Las líneas punteadas son rectas  $q\alpha$  para  $q = -10, \dots, 10$ . El punto de tangencia (mínima distancia vertical) define  $\alpha_q$ .

La función  $\tau(q)$  obtenida mediante esta transformación se muestra en la Figura 3.27. Nótese que  $\tau(q)$  es cóncava y sus asíntotas para  $q \rightarrow \pm\infty$  son las rectas  $q\alpha_{\min}$  y  $q\alpha_{\max}$ . La derivada satisface  $\frac{d\tau}{dq} = \alpha_q$ , de acuerdo con la teoría de la transformada de Legendre. Recíprocamente, a partir de  $\tau(q)$  puede recuperarse  $f(\alpha)$  mediante  $f(\alpha) = \inf_q \{q\alpha - \tau(q)\}$ .



**Figura 3.27:** Función  $\tau(q)$  para la medida de Cantor (curva sólida). Las líneas diagonales discontinuas son las asíntotas  $q\alpha_{\min}$  y  $q\alpha_{\max}$ . Las líneas punteadas verticales ilustran la distancia  $f(\alpha_q) = q\alpha_q - \tau(q)$ .

## 3.7. Resumen y Conclusiones del Capítulo

En este capítulo se ha desarrollado el marco teórico de la geometría multifractal, partiendo de ejemplos concretos (como el conjunto de Cantor y la bipartición binaria) hasta llegar a una formulación general basada en medidas multinomiales. Se han introducido y relacionado dos descripciones complementarias: la local, mediante el espectro  $f(\alpha)$  que da la dimensión de los conjuntos de puntos con singularidad  $\alpha$ , y la global, mediante la familia de dimensiones generalizadas  $D_q$  o su equivalente  $\tau(q)$ .

Se ha destacado el origen informático de estos conceptos, remontándose a las entropías de Rényi, y se ha mostrado su utilidad en el análisis de sistemas complejos como la turbulencia completamente desarrollada y los atractores caóticos. Un resultado central es el teorema de Eggleston, que revela la naturaleza multifractal del propio intervalo unitario cuando se le dota de una medida no uniforme.

La dualidad entre las descripciones local y global se cristaliza en la transformada de Legendre, que conecta  $f(\alpha)$  y  $\tau(q)$ . Esta relación no es solo formal; establece una profunda analogía con la mecánica estadística, donde  $\alpha$  juega el papel de energía,  $q$  el de temperatura inversa,  $\tau$  el de potencial termodinámico y  $f$  el de entropía. Así, el formalismo multifractal proporciona un puente poderoso entre la geometría fractal, la teoría de la información y la física estadística.

## Parte II

# Aplicación al Genoma Humano



---

El análisis computacional de secuencias genómicas enfrenta actualmente desafíos fundamentales derivados de la escala masiva de los datos genómicos modernos. Con el genoma humano completo T2T-CHM13v2.0 ocupando aproximadamente 3.1Gb de información, los métodos tradicionales de representación y análisis muestran limitaciones significativas en precisión, escalabilidad y capacidad de caracterización estructural. Esta segunda parte de la tesis presenta una contribución metodológica integral que aborda estos desafíos mediante el desarrollo de dos enfoques complementarios: la Representación Binaria Genómica como reformulación fundamental del Chaos Game Representation, y el análisis multifractal como marco para la caracterización cuantitativa de la organización genómica.

El Capítulo 4 establece los fundamentos teóricos y computacionales de la Representación Binaria Genómica (BGR), partiendo de una revisión crítica del Chaos Game Representation tradicional. El capítulo comienza examinando el Juego del Caos de Barnsley para generación de fractales mediante sistemas de funciones iteradas, destacando las adaptaciones específicas que Jeffrey introdujo para su aplicación a secuencias genómicas.

La limitación central identificada en el enfoque tradicional radica en la pérdida progresiva de precisión numérica y la incapacidad para reconstruir completamente la secuencia genómica a partir de los puntos gráficos. La BGR emerge como solución a estas limitaciones mediante una visión fundamental: la expresión analítica de las reglas iterativas del CGR revela que la representación binaria constituye la base natural para el estudio de secuencias genómicas.

El desarrollo teórico procede mediante la asociación de cada nucleótido con vectores binarios bi-dimensionales, transformando así la secuencia genómica completa en un par de secuencias binarias. Esta reformulación permite reinterpretar el CGR como la trayectoria exacta de un caminante determinista, donde cada punto de la trayectoria encapsula toda la información histórica de la secuencia. Se demuestra formalmente que el último punto de la trayectoria contiene información completa sobre toda la secuencia genómica, permitiendo la reconstrucción exacta mediante relaciones de recurrencia inversa.

La introducción del parámetro  $M$  constituye una innovación crucial, proporcionando control explícito sobre la resolución y escala de las representaciones mediante el Caminante Determinista Compactado. Este enfoque genera representaciones gráficas en espacios de  $2^M \times 2^M$  puntos, superando las limitaciones de precisión numérica del CGR tradicional. La equivalencia formal establecida entre los esquemas de Jeffrey y Yin a través del formalismo BGR permite no solo representaciones visuales exactas, sino también esquemas de codificación eficientes mediante vectores enteros.

El capítulo explora exhaustivamente las propiedades geométricas del espacio CGR mediante el operador de Hutchinson, mostrando cómo el cuadrado unitario se divide recursivamente en regiones correspondientes a subsecuencias específicas. Esta estructura fractal proporciona el fundamento matemático para la localización exacta de puntos y el análisis de frecuencias de subsecuencias. Las aplicaciones prácticas desarrolladas incluyen algoritmos eficientes para codificación y decodificación,

---

métodos para el análisis de frecuencias de subsecuencias, y técnicas para la identificación de patrones estructurales en secuencias genómicas largas.

El Capítulo 5 traslada el marco metodológico desarrollado al dominio del análisis cuantitativo mediante la aplicación exhaustiva del formalismo multifractal al genoma humano completo. El capítulo comienza estableciendo el contexto biológico esencial, revisando los principios de citogenética humana y las características estructurales fundamentales de los cromosomas, incluyendo la organización en bandas, la localización de centrómeros, y la distribución de regiones funcionales como genes, promotores, y elementos reguladores.

La aplicación del análisis multifractal se implementa mediante técnicas de box-counting sistemáticas, calculando espectros completos de singularidad, exponentes de masa, y dimensiones generalizadas para el genoma completo y cada cromosoma individual. El análisis revela que la organización estructural del genoma humano exhibe heterogeneidades complejas que requieren caracterización más allá de las dimensiones fractales tradicionales.

Los resultados demuestran patrones multifractales distintivos entre cromosomas, con la mayoría exhibiendo espectros similares centrados alrededor de dimensiones fractales consistentes, pero con variaciones significativas en sus características específicas. Los cromosomas 9 e Y emergen como casos particulares, mostrando los mayores rangos de exponentes de singularidad y comportamientos anómalos en sus espectros multifractales.

El capítulo desarrolla una comparación sistemática entre las representaciones BGR y de Cadenas de Markov, demostrando su complementariedad esencial. Mientras la BGR captura más efectivamente componentes de alta frecuencia asociados con regiones repetitivas, las Cadenas de Markov se alinean mejor con regiones de baja frecuencia ricas en dinucleótidos CG. Para cadenas de doce nucleótidos, el enfoque de Cadenas de Markov logra un error promedio de sólo 2 % respecto a la representación completa.

La aplicación específica a los cromosomas 9 e Y ilustra el poder de estos métodos para identificar características estructurales específicas. El análisis identifica secuencias altamente repetitivas en regiones heterocromáticas, revela patrones de bandas citogenéticas mediante transformaciones unidimensionales del CGR, y localiza precisamente polimorfismos y variaciones estructurales. Estas capacidades trascienden el análisis meramente descriptivo, proporcionando herramientas cuantitativas para la caracterización de arquitecturas genómicas específicas.

La integración metodológica final establece un proceso completo que combina representaciones exactas mediante BGR con caracterización cuantitativa mediante análisis multifractal, ofreciendo un marco unificado para el estudio de la organización estructural de secuencias genómicas a múltiples escalas. Este enfoque integrado no sólo supera limitaciones metodológicas previas, sino que establece las bases para aplicaciones en genómica comparada, identificación de variantes estructurales, y estudios de evolución genómica a gran escala.



# Capítulo 4

## Representación Genómica Binaria

### 4.1. Generación Fractal con el Juego del Caos

Para generar fractales gráficamente, [2] construyó un caminante aleatorio donde los puntos visitados convergen rápidamente a puntos del fractal, recorriendo aleatoriamente diferentes regiones del mismo. Para construir el caminante aleatorio, se introduce un sistema de funciones iteradas contractivas (IFS), donde se aplican iterativamente funciones seleccionadas aleatoriamente a un punto arbitrario  $P_0$ . Esto se logra utilizando un sistema probabilístico finito [45], donde los eventos son  $n$  funciones contractivas  $w_1(x), w_2(x), \dots, w_n(x)$  en un espacio métrico  $X$ , junto con sus probabilidades  $p_1, p_2, \dots, p_n$ , es decir,

$$S = \begin{pmatrix} w_1(x) & w_2(x) & \dots & w_n(x) \\ p_1 & p_2 & \dots & p_n \end{pmatrix}, \quad (4.1)$$

satisfaciendo:

$$\left\{ p_i \geq 0, \sum_{i=1}^n p_i = 1 \right\}$$

donde las funciones son contractivas, es decir,

$$d(w_i(x_1), w_i(x_2)) < d(x_1, x_2) \quad \forall i \in \{1, 2, \dots, n\}, \quad (4.2)$$

La trayectoria del caminante aleatorio está dada por  $P_{j+1} = w_j(P_j)$ , donde  $j$  se elige aleatoriamente del conjunto  $\{1, 2, \dots, n\}$  con probabilidades  $\{p_1, p_2, \dots, p_n\}$  y  $P_0 \in X$  es un punto arbitrario. Los puntos generados no son aleatorios sino que forman imágenes fractales que corresponden a la representación gráfica de la medida estadística estacionaria asociada con el IFS. Este método de generación fractal se conoce como el Juego del Caos de Barnsley (BCG) [2].

Para obtener una representación gráfica del genoma, [6], inspirado por el BCG, introdujo la Representación del Juego del Caos (CGR), donde reemplaza el componente estocástico del BCG con una secuencia determinista correspondiente a una Secuencia Genómica (SG). El resultado es sorpren-

dente, ya que revela que cada SG tiene un fractal único asociado, permitiendo obtener gráficamente la “huella digital” de los genomas de organismos vivos [21].

#### 4.1.1. Representación gráfica del ADN

Para obtener una representación gráfica del ADN, Jeffrey construye un cuadrado unitario  $Q$ , asociando cada uno de sus vértices con una de las cuatro bases que componen el alfabeto de la secuencia genómica, denotadas por las letras A, C, G y T. Así, los vértices de  $Q$  son:

$$A = \begin{pmatrix} 0 \\ 0 \end{pmatrix}, C = \begin{pmatrix} 0 \\ 1 \end{pmatrix}, G = \begin{pmatrix} 1 \\ 1 \end{pmatrix}, T = \begin{pmatrix} 1 \\ 0 \end{pmatrix}; \quad (4.3)$$

Dada la SG de  $N$  nucleótidos:

$$V_1 V_2 \dots V_N = \begin{pmatrix} V_{1x} V_{2x} \dots V_{Nx} \\ V_{1y} V_{2y} \dots V_{Ny} \end{pmatrix}; \quad V_i \in \{A, C, G, T\}, \quad (4.4)$$

asociamos  $N$  puntos dentro del cuadrado unitario  $Q$  usando el siguiente proceso iterativo:

- (a) Seleccionar  $P_0$ , el punto medio de  $Q$ .
- (b) Asociar el primer término de (4.4),  $V_1$ , con el punto  $P_1$ , que es el punto medio del segmento que conecta  $P_0$  con el vértice del cuadrado correspondiente al primer nucleótido de la secuencia genómica.
- (c) Asociar el segundo término de la SG,  $V_2$ , con el punto  $P_2$ , que es el punto medio del segmento que conecta el punto  $P_1$  con el vértice de  $Q$  correspondiente al segundo nucleótido de la secuencia.
- (d) Repitiendo este proceso para toda la secuencia de ADN se generan tantos puntos como nucleótidos en la secuencia de ADN dada. (Ver Figura 4.1)

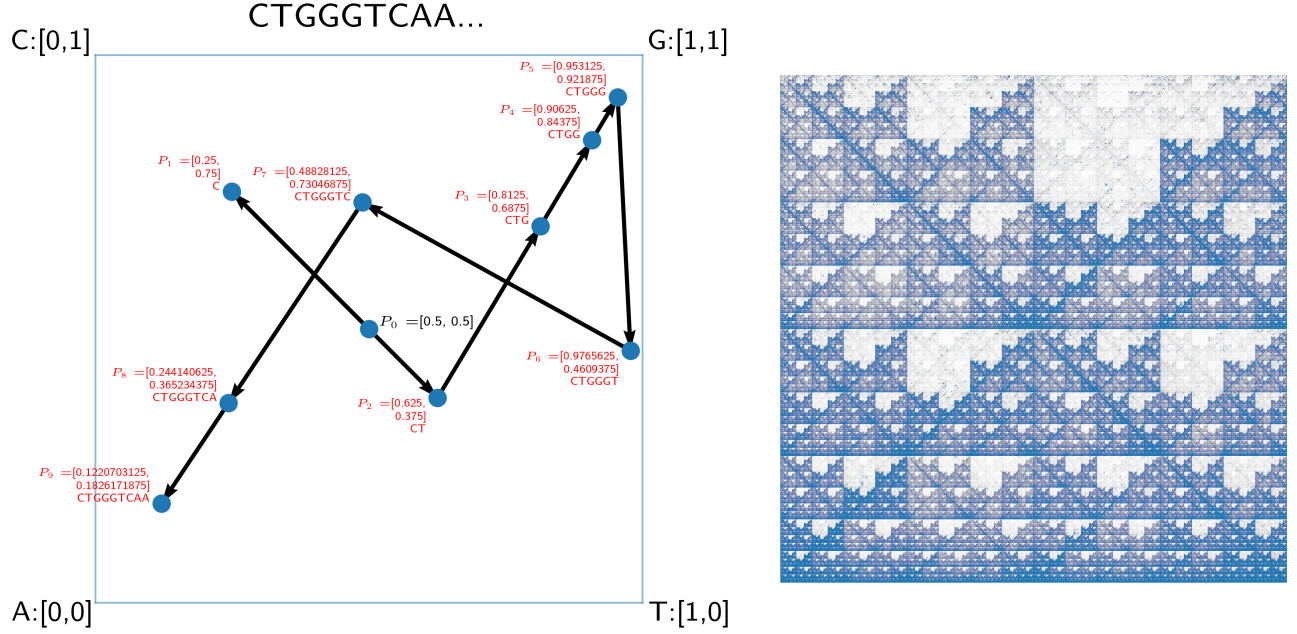
Para expresar analíticamente el proceso anterior, Jeffrey utiliza el formalismo IFS introducido por Hutchinson y Barnsley [2, 26] para construir gráficos de objetos fractales, definiendo el IFS con las siguientes cuatro funciones contractivas:

$$w_i(\mathbf{X}) = \frac{1}{2} \left( \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} + \begin{pmatrix} V_{ix} \\ V_{iy} \end{pmatrix} \right); \quad i \in \{1, 2, 3, 4\}, \quad (4.5)$$

donde  $\mathbf{X} \in \mathbb{R}^2$  y los puntos fijos de estas funciones corresponden a los vértices de  $Q$ .

Dada la SG (4.4), los puntos  $P_1, P_2, \dots, P_N$  se asignan dentro de  $Q$  usando la regla de iteración:

$$P_{n+1} = w_i(P_n) = \frac{1}{2} (P_n + V_{n+1}), \quad (4.6)$$



**Figura 4.1:** El panel izquierdo muestra la Representación del Juego del Caos de Jeffrey para una secuencia ejemplo, mientras que el panel derecho muestra el CGR completo del cromosoma 1 humano. A medida que se procesa la secuencia, el número requerido de dígitos decimales aumenta en uno. La precisión máxima utilizada en la representación del cromosoma 1 es de 17 dígitos decimales, lo que significa que cada punto codifica solo los 15 nucleótidos precedentes. El patrón resultante exhibe comportamiento autosimilar.

para  $n \geq 1$  y  $P_0 = V_0 = (0.5, 0.5)^T$ . Graficando estos puntos se produce la representación gráfica de la SG.

Dado que la metodología CGR solo requiere almacenar un punto para generar el siguiente, el algoritmo es altamente eficiente, con resultados mostrados directamente en el monitor de la computadora. Sin embargo, generalmente permite solo análisis gráfico, lo que limita su aplicación a SG muy grandes, donde se emplean varios métodos estadísticos [22], o para analizar cromosomas específicos dentro de secuencias de ADN [21]. Esta metodología restringe el estudio analítico de secuencias, ya que almacenar las coordenadas de todos los puntos requiere una gran cantidad de memoria. La Representación Genómica Binaria (BGR) supera esta limitación, permitiendo estudios analíticos de todo el genoma.

Una comparación entre CGR y BCG revela que ambos usan un IFS compuesto de funciones contractivas pero difieren en su elección del punto inicial y la manera en que se aplica el IFS. En BCG, el punto de partida puede ser cualquier ubicación dentro de los vértices del cuadrado, mientras que en CGR siempre se selecciona el punto medio del cuadrado. La segunda distinción radica en cómo se aplica el IFS: BCG emplea un enfoque estocástico, aplicando el IFS aleatoriamente en cada iteración, mientras que CGR lo aplica determinísticamente, guiado por el orden y tipo de nucleótidos en la SG. Además, los puntos que definen la trayectoria en CGR son finitos, correspondiendo al número de nucleótidos en la SG, creando así un Caminante Determinista (DW). Esta naturaleza

determinista asegura que cualquier punto a lo largo de la trayectoria del DW codifique suficiente información para reconstruir tanto los puntos precedentes como la secuencia de nucleótidos que los generó.

Una forma simple de demostrar esta propiedad es invirtiendo (4.6), usando el hecho de que cada punto en la trayectoria del DW está dentro del cuadrado unitario  $Q$  formado por los vértices  $(V_A, V_C, V_G, V_T)$ , obteniendo el siguiente resultado [16–18]:

$$P_{n-1} = 2P_n \quad \text{mód } 1 \quad \text{y} \quad V_n = \lfloor 2P_n \rfloor, \quad (4.7)$$

con  $\lfloor b \rfloor$  representando la parte entera de  $b$ , conociendo  $P_n$  podemos encontrar  $P_{n-1}$  y  $V_n$ . En principio, este resultado puede usarse iterativamente para determinar los primeros  $n-1$  puntos en la trayectoria del caminante  $(P_1, P_2, \dots, P_{n-1})$  y los primeros  $n$  nucleótidos formando la SG:  $(V_1, V_2, \dots, V_n)$ . Este enfoque es válido para cualquier valor de  $n$ , lo que significa que si  $P_N$ , el último punto en la SG, es conocido, todos los puntos visitados por el DW y la SG pueden reconstruirse.

Sin embargo, este enfoque requiere las coordenadas precisas de  $P_n$ , lo cual, en la práctica, no es posible, ya que los puntos gráficos proporcionan solo una posición aproximada de cualquier punto en la trayectoria del DW. Esto implica que solo un número limitado de puntos precedentes a  $P_n$  puede determinarse, incluso con varios análisis gráficos y estadísticos [22]. No obstante, tales análisis no permiten reconstruir la SG completa usando solo un punto de la trayectoria del DW.

## 4.2. Representación Genómica Binaria (BGR)

Una forma de aplicar (4.7) es usar una Representación Genómica Binaria (BGR) dentro del marco del CGR. Esto implica reexpresar la ubicación de puntos dentro de  $Q$ , como se da en (4.6), en términos de los nucleótidos en la SG, obteniendo el punto asociado con la SG en (4.4):

$$\vec{P}_R = \frac{\vec{V}_R}{2} + \frac{\vec{V}_{R-1}}{2^2} + \dots + \frac{\vec{V}_1 + \vec{P}_0}{2^R} = \sum_{j=0}^R \frac{\vec{V}_j}{2^{R-j+1}}; \quad (4.8)$$

con  $\vec{V}_0 = (1, 1)^T$ . Esta relación resalta la conveniencia de usar una base binaria. Al expresar los vectores correspondientes al centro y vértices de  $Q$  en coordenadas binarias (ver (4.3)), encontramos las coordenadas de  $P_N$  en binario:

$$\begin{aligned} \vec{P}_N &= \sum_{j=0}^N \frac{\vec{V}_j}{2^{N-j+1}} = \vec{0} \cdot \vec{V}_N \vec{V}_{N-1} \dots \vec{V}_1 \vec{V}_0 \\ &= \begin{pmatrix} P_{N_X} \\ P_{N_Y} \end{pmatrix} = \begin{pmatrix} 0.V_{N_x} V_{(N-1)_x} \dots V_{1_x} V_{0_x} \\ 0.V_{N_y} V_{(N-1)_y} \dots V_{1_y} V_{0_y} \end{pmatrix}. \end{aligned} \quad (4.9)$$

Este resultado, combinado con (4.3), permite identificar el nucleótido en la SG en la posición  $N$  como  $\vec{V}_N = \lfloor 2\vec{P}_N \rfloor$ , y permite encontrar todos los puntos  $P_S$  con  $S < N$ , así como identificar el nucleótido en la posición  $S$  en la SG, es decir,

$$\vec{P}_S = \text{Frac}\left(2^{N-S}\vec{P}_N\right); \quad \vec{V}_S = \lfloor 2\vec{P}_S \rfloor, \quad (4.10)$$

donde  $\text{Frac}(x) = x - \lfloor x \rfloor$  representa la parte fraccionaria.

De esta manera, BGR demuestra que, conociendo el número de nucleótidos  $N$  en la SG y las coordenadas de  $P_N$ , el último punto en la SG, se pueden determinar las coordenadas de todos los puntos precedentes e identificar los nucleótidos que representan usando (4.3).

Al leer la SG secuencialmente, BGR permite almacenar dos secuencias de 0s y 1s. Para secuencias grandes, puede usarse un proceso de concatenación para almacenarlas, obteniendo una representación numérica de la SG completa. Esto permite el análisis de SG con un mayor número de nucleótidos de lo que es posible usando el método estándar.

Los puntos generados con CGR en binario son:

$$O(\vec{P}_0) = \left\{ 0.\vec{V}_0, 0.\vec{V}_1\vec{V}_0, 0.\vec{V}_2\vec{V}_1\vec{V}_0, \dots, 0.\vec{V}_N \dots \vec{V}_2\vec{V}_1\vec{V}_0 \right\} \quad (4.11)$$

mostrando que estos puntos corresponden a la órbita generada por el punto medio de  $Q$ :  $P_0 = 0.V_0 = (0.1, 0.1)^T$ , cuando el IFS dado en (4.5) se aplica en el orden en que se lee la SG. Este enfoque permite ubicar los puntos generados por CGR expresando la órbita  $O(P_0)$  en términos del IFS. La órbita generada por la SG dada en (4.4) es:

$$O(\vec{P}_0) = \left\{ \vec{P}_0, w_{V_1}(\vec{P}_0), w_{V_2} \circ w_{V_1}(\vec{P}_0), \dots, w_{V_N} \circ \dots \circ w_{V_2} \circ w_{V_1}(\vec{P}_0) \right\}. \quad (4.12)$$

Esta expresión permite ubicar todos los puntos generados por la SG, aproximar los puntos gráficos y calcular la frecuencia de visitas de la órbita a través de los diversos cubrimientos del atractor de la SG, un paso necesario para el análisis multifractal.

En resumen, al expresar la SG completa en binario, puede almacenarse fácilmente, y se observa que el último punto de la órbita contiene toda la información sobre sus subsecuencias. Así, para ubicar la posición del  $R$ -ésimo punto en  $O(\vec{V}_0)$ , basta con desplazar el punto decimal  $N - R$  lugares a la derecha y considerar la parte fraccionaria tanto para las coordenadas  $x$  como  $y$ , es decir,

$$\begin{aligned} X_R &= 0.V_{R_x}V_{(R-1)_x} \dots V_{1_x}V_{0_x}; \\ Y_R &= 0.V_{R_y}V_{(R-1)_y} \dots V_{1_y}V_{0_y}. \end{aligned} \quad (4.13)$$

Así, al representar la SG completa en binario, el punto final de la órbita (4.11) puede almacenarse en la memoria de la computadora, y con un simple apuntador, pueden recuperarse las coordenadas

de cualquier punto asociado con una subsecuencia sin ningún cálculo, ya que el último punto de la órbita contiene toda la información sobre sus subsecuencias.

La misma información contenida en  $\vec{P}_N \in Q$  está dada en el vector entero:

$$\begin{aligned} \vec{Y}_N &= \sum_{j=0}^N 2^j \vec{V}_j = \begin{pmatrix} V_{N_x} V_{(N-1)_x} \cdots V_{1_x} V_{0_x} \\ V_{N_y} V_{(N-1)_y} \cdots V_{1_y} V_{0_y} \end{pmatrix} \\ &= 2^{N+1} \vec{P}_N. \end{aligned} \quad (4.14)$$

Este se encuentra fuera de  $Q$  y da el esquema alternativo del CGR propuesto por Yin [16] llamado Representación del Juego del Caos Entera (iCGR), donde toda la SG se caracteriza por tres números enteros: el número  $N + 1$  de nucleótidos (incluyendo el pseudo-nucleótido  $V_0$ ) que conforman la SG, y las coordenadas enteras asociadas con su último nucleótido, dadas por (4.14).

La ventaja de usar BGR es que muestra claramente la equivalencia entre los esquemas de Jeffrey y Yin del CGR, debido a que existe una correspondencia uno a uno entre los puntos  $P_R$  dentro de  $Q$  con el punto  $Y_R$  fuera de  $Q$ .

El esquema de Yin no tiene una representación gráfica simple, pero el estudio analítico de la SG es más simple que el esquema de Jeffrey; de hecho, el esquema de Yin sería muy útil para codificar y decodificar SG muy grandes como vemos en la sección 4.3, y usando este procedimiento de compactación es posible obtener una representación gráfica (GR). La GR se obtiene introduciendo un Caminante Determinista Compactado (CDW) donde cada punto codifica la información de un grupo de  $M$  nucleótidos, y como este CDW está conformado por  $p + 1$  puntos menos que  $N$ , es posible obtener la imagen gráfica de la SG en el esquema de Yin.

La imagen gráfica para SG grandes se obtiene usando BGR en el iCGR y el CDW donde cada punto codifica la información de un grupo de  $M$  nucleótidos. La representación gráfica de este CDW corresponde con la imagen gráfica de la SG.

#### 4.2.1. Ubicación de puntos en la órbita del $DW$ y sus puntos gráficos

Para ubicar los puntos de  $O(\vec{V}_0)$ , utilizamos el Teorema de Hutchinson [26], el cual establece que dado un IFS que consiste en  $n$  funciones contractivas

$$\{X; w_1(x), w_2(x), \dots, w_n(x)\},$$

podemos definir el operador de Hutchinson  $W$ :

$$W(A) = \bigcup_{i=1}^n w_i(A). \quad (4.15)$$

Cuando se aplica iterativamente a un conjunto  $A \subset X \in \mathbb{R}^2$ , este operador produce el fractal  $F$  asociado con el IFS:

$$F = \lim_{k \rightarrow \infty} W^n(A); \quad \forall A \in \mathbb{R}^2 \quad (4.16)$$

implicando que el punto fijo de (4.15) es un conjunto único  $F$  generado por el IFS, tal que:

$$W(F) = F. \quad (4.17)$$

Cuando  $w_i(F) \cap w_j(F) = \emptyset$  si  $i \neq j$ , aplicar  $W$  a  $F$   $P$  veces separa el conjunto en  $N^P$  regiones distintas, permitiéndonos identificar  $C_P$ , el cubrimiento óptimo de orden  $P$  para  $F$ , es decir,

$$C_P = W^P(F) = W \circ W \circ \dots \circ W(F); \quad (4.18)$$

donde  $F \subset C_P \quad \forall P \in \mathbb{Z}^+$ .

Usando el Teorema de Hutchinson, analizamos el operador  $W$  asociado con el CGR. Para el IFS en (4.5),  $W$  se define como:

$$W(R) = w_A(R) \cup w_C(R) \cup w_G(R) \cup w_T(R); \quad \forall R \in \mathbb{R}^2, \quad (4.19)$$

que, cuando se aplica al cuadrado unitario  $Q$ , resulta en:

$$W(Q) = Q_A \cup Q_C \cup Q_G \cup Q_T,$$

donde

$$\begin{aligned} Q_A &= W_A(Q), & Q_C &= W_C(Q), \\ Q_G &= W_G(Q), & Q_T &= W_T(Q), \end{aligned} \quad (4.20)$$

Así,  $W$  divide  $Q$  en cuatro cuadrados con longitud de lado  $L_1 = \frac{1}{2}$ , cuya unión forma  $Q$ . Por lo tanto, el atractor generado por (4.5) es el cuadrado unitario abierto, es decir,

$$F = \{(x, y) | x, y \in (0, 1)\}. \quad (4.21)$$

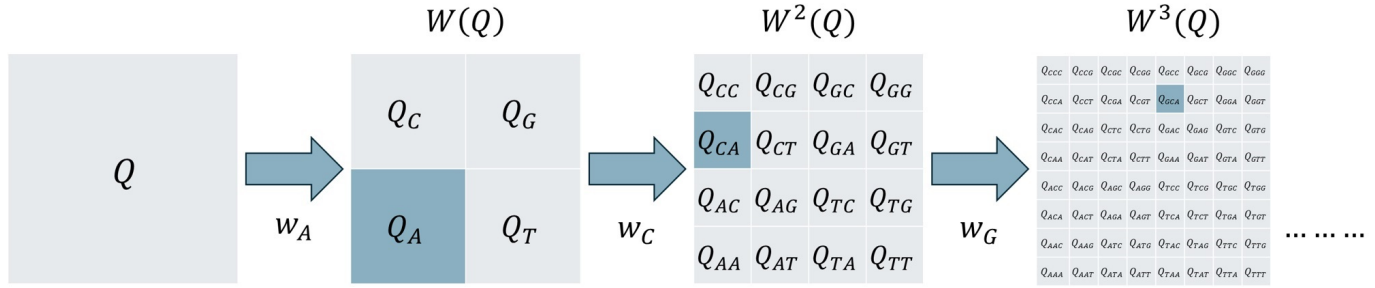
Es directo mostrar que las dimensiones de caja y Hausdorff de  $F$  son ambas iguales a 2. Este conjunto sirve como soporte para la medida estadística generada por cualquier SG, y el conjunto atractor de la SG está contenido dentro de  $F$ .

Cuando  $W$  se aplica dos veces a  $Q$ , encontramos:

$$W^2(Q) = Q = \bigcup_{V_2=A}^T \bigcup_{V_1=A}^T Q_{V_2V_1}, \quad (4.22)$$

lo cual divide  $Q$  en  $4^2$  cuadrados con longitud de lado  $L_2 = \left(\frac{1}{2}\right)^2$ , como se muestra en la Figura 4.2. Al aplicar  $W$   $n$  veces a  $Q$ , este se divide en  $4^n$  cuadrados con longitud de lado  $L_n = \left(\frac{1}{2}\right)^n$ , lo que forma un cubrimiento  $C_n$  de  $Q$ . La familia contable  $\{C_n\}$  de estos cubrimientos se llama la estructura fractal finita de  $F$  [46].

Las ubicaciones de los puntos en la órbita generada por  $\vec{V}_0$ , el punto medio (MP) de  $Q$  (ver Figuras 4.1 y 4.2), están dadas por:  $\vec{P}_1 = w_{V_1}(\vec{P}_0)$ , el MP de  $Q_{V_1}$ ;  $\vec{P}_2 = w_{V_2} \circ w_{V_1}(\vec{P}_0)$ , el MP de



**Figura 4.2:** Transformaciones  $W(Q)$ ,  $W^2(Q)$  y  $W^3(Q)$  con la ubicación de sus componentes. Estas transformaciones revelan la localización de regiones asociadas con secuencias que terminan en subsecuencias específicas, asignando una región distinta a cada posible terminación de longitud  $M$ .

$Q_{V_2V_1}$ ; y  $\vec{P}_n = w_{V_n} \circ w_{V_{n-1}} \circ \dots \circ w_{V_1}(\vec{P}_0)$ , el MP de  $Q_{V_nV_{n-1}\dots V_1}$ . Las ubicaciones de los puntos del DW muestran que forman un conjunto finito y desconectado, que en el límite cuando  $N \rightarrow \infty$  se convierte en un conjunto de tipo Cantor.

Para establecer la conexión entre el DW y el análisis gráfico, partimos de la órbita completa  $O(\vec{P}_0)$  dada en (4.12) y asumimos que los píxeles son cuadrados pequeños de tamaño  $2^{-K} \times 2^{-K}$ , correspondiendo a un cubrimiento  $C_K$ . Así, solo pueden distinguirse puntos aproximados dentro de  $Q$ , con coordenadas binarias  $\tilde{P}_g = (X = 0.a_1 \dots a_K, Y = 0.b_1 \dots b_K)$ , y solo los primeros  $K$  puntos de  $O(\vec{P}_0)$  pueden graficarse con precisión. Más allá de esto, solo los primeros  $K$  dígitos de cada punto en la órbita son gráficamente relevantes, por lo que cualquier punto  $\tilde{P}_R$  se aproxima a  $K$  dígitos significativos, es decir,

$$\begin{aligned} \vec{P}_R &= 0.\vec{V}_R\vec{V}_{R-1}\dots\vec{V}_2\vec{V}_1\vec{V}_0 \\ &\rightarrow \tilde{P}_R \approx 0.\vec{V}_R\vec{V}_{R-1}\dots\vec{V}_{R-K+2}\vec{V}_{R-K+1}. \end{aligned} \quad (4.23)$$

Así, el DW se comporta de manera similar a una partícula browniana moviéndose en un líquido, siendo golpeada por innumerables moléculas sin la capacidad de observar cada impacto individual. El gráfico de la órbita proporciona una aproximación gruesa de la órbita exacta. Sin embargo, a diferencia de una partícula browniana, la órbita exacta en el CGR es conocida, permitiéndonos determinar con precisión cuántos puntos en la órbita caen dentro de cada píxel. Esto permite un proceso de compactación versátil y eficiente para graficar la SG completa, como se mostrará en la siguiente sección, y también respalda varios análisis estadísticos, particularmente un estudio multifractal del genoma.

Es posible construir un esquema de cubrimiento óptimo que evolucione con cada iteración del CGR, considerando la unión de todas las etapas anteriores. Este enfoque nos permite visualizar no solo los resultados posibles en la iteración  $n$ , sino también los de todas las secuencias anteriores. Dado que la ubicación de los puntos cambia con cada iteración—como se observa en el caso del cuadrado unitario—cada paso en el CGR efectivamente añade un dígito decimal a cada secuencia posible.



### Espacio CGR

El conjunto de puntos en  $\mathbb{R}^2$  que contiene todos los CGRs posibles de longitud  $L \leq N$  puede por lo tanto describirse como:

$$\mathbf{X} = \bigcup_{n=1}^N \{2^{-n-1} (2i+1, 2j+1) \mid i, j \in \mathbb{N}, 0 \leq i, j \leq 2^n - 1\}, \quad (4.24)$$

o aplicando el operador de Hutchinson al punto  $P_0$ :

$$\mathbf{X} = \bigcup_{n=1}^N \mathbf{W}^n(P_0).$$

Este conjunto corresponde a la unión de mallas centradas en subcuadrados de una malla uniforme sobre  $[0, 1]^2$  con resolución  $2^{-n}$ . El número total de cubrimientos está dado por  $\sum_{i=1}^N 4^i$ , que en el límite cuando  $N \rightarrow \infty$  llena densamente el cuadrado unitario. Por lo tanto, el número total de píxeles necesarios en una pantalla de monitor para mostrar fielmente toda la SG corresponde a  $(2^{N+1} - 1) \times (2^{N+1} - 1)$ .

A diferencia del operador de Hutchinson, que produce cubrimientos para cualquier secuencia que termina en una subsecuencia finita de orden  $n$ , esta construcción define cubrimientos óptimos por punto hasta un orden específico. En consecuencia, los cubrimientos de Hutchinson necesariamente contienen todos los puntos de este conjunto.

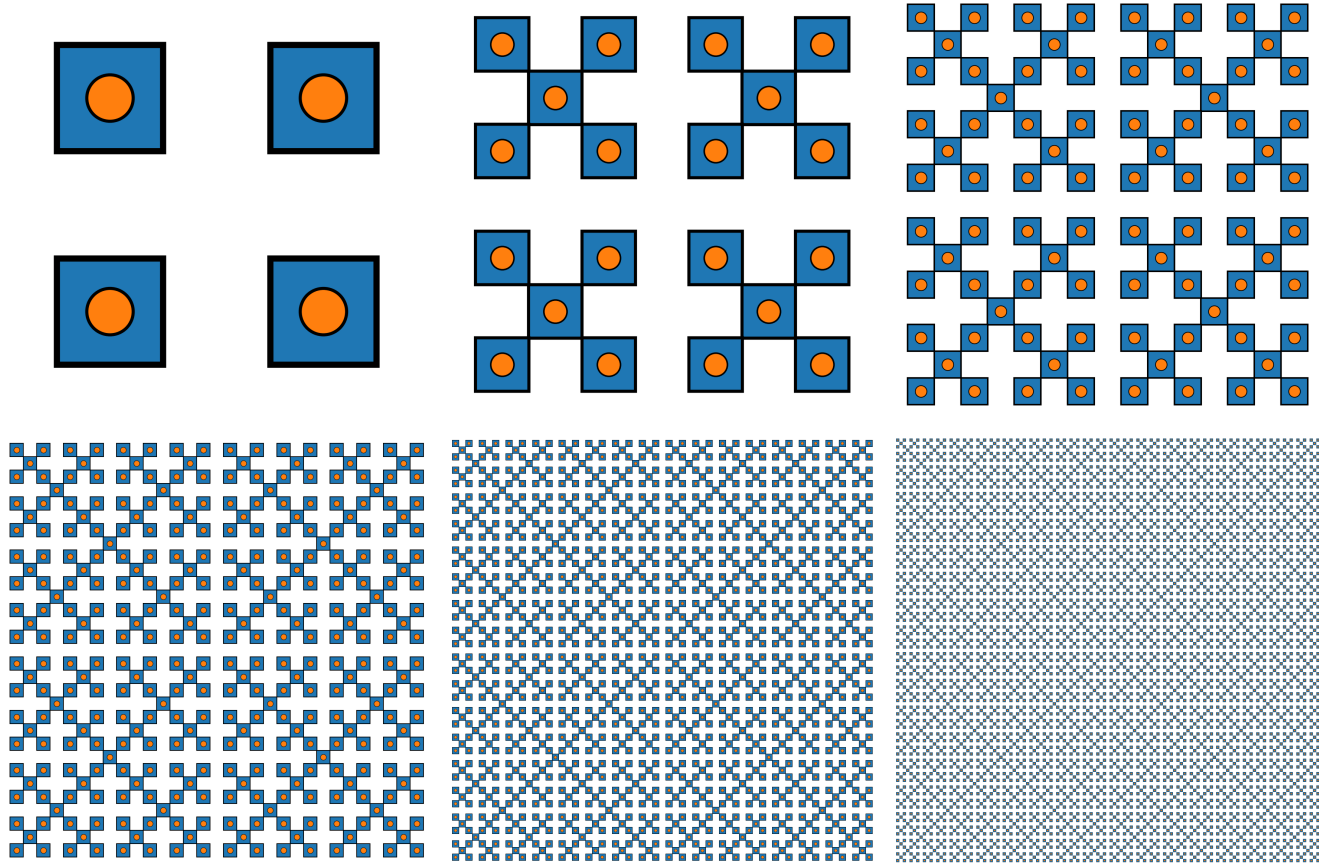
Un ejemplo que ilustra todos los cubrimientos posibles para CGRs hasta el orden  $N$  se muestra en la Figura 4.3, donde cada punto corresponde a una secuencia específica.

### Espacio BGR

La distribución de puntos en el CGR difiere significativamente de la de las representaciones binarias, que reproducen solo subsecuencias de longitud fija de tamaño  $M$ . Esto implica que el espacio de representación binaria es un subconjunto estricto del espacio CGR, ya que está restringido a una malla uniformemente espaciada sobre el dominio  $[0, 2^M - 1]^2$  (ver Figura 4.4). Esta restricción limita la representación a las frecuencias asociadas con cada punto—es decir, la frecuencia de ocurrencia para cada subsecuencia de longitud  $M$ .

Otra característica clave de la BGR es que, a partir de la representación en un  $M$  dado, es posible derivar las representaciones para todos los valores menores de  $M$ . Aprovechando la organización jerárquica definida por la BGR, la representación de tamaño  $M-1$  se obtiene sumando las frecuencias sobre bloques adyacentes y reescalando el dominio al espacio de puntos  $[0, 2^{M-1} - 1]^2$ .

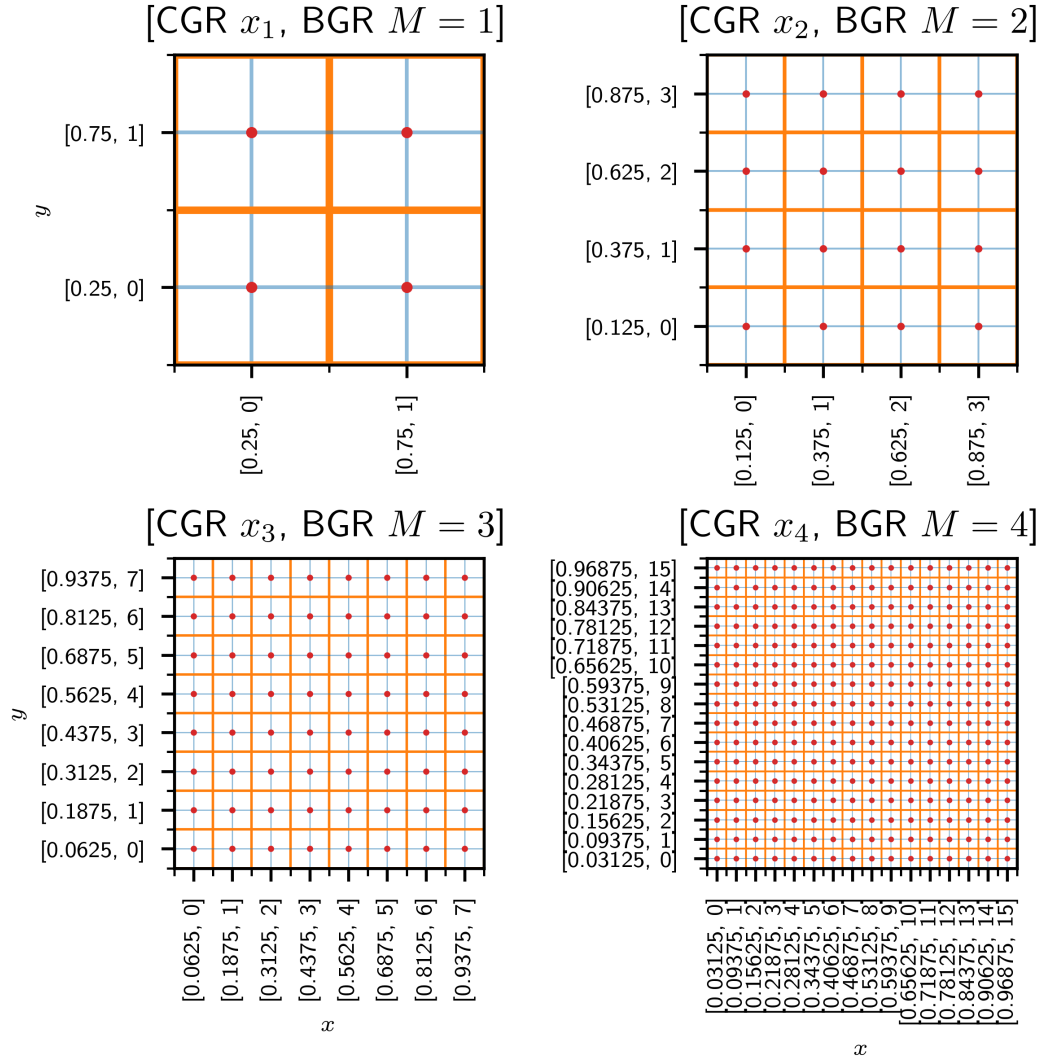
No obstante, la BGR ofrece un método de codificación rápido y representación gráfica de distribuciones CGR de secuencias de tamaño  $M$ . Dada la equivalencia punto por punto entre las dos representaciones, la unión de BGRs para  $1 \leq M \leq N$ , cada una reescalada por un factor de  $\frac{2^N - 1}{2^M - 1}$ , se vuelve equivalente al espacio CGR.



**Figura 4.3:** Conjunto de cubrimientos para todos los puntos CGR posibles correspondientes a secuencias de longitudes máximas desde 1 hasta 6. Esto constituye una agregación fractal [3] que llena  $\mathbb{R}^2$  de acuerdo con la Ecuación 4.24.

Finalmente, considerando que la precisión numérica en computadoras digitales es inherentemente limitada, el espacio de puntos CGR representables está acotado por el número de dígitos que pueden almacenarse. Además, almacenar todos los puntos del CGR hasta longitud  $N$  requeriría un total de  $\sum_{n=1}^N n$  dígitos. Por lo tanto, puede ser más eficiente almacenar solo el valor BGR final para una secuencia de longitud  $M$ , o aplicar la estrategia de decodificación descrita en la siguiente sección.

La Figura 4.5 muestra una imagen obtenida usando la BGR con  $M = 20$  para la secuencia genómica humana completa, incluyendo ADN mitocondrial. Para producir una imagen discernible, los puntos se difuminaron; de lo contrario, el resultado aparecería como un esquema casi homogéneo con vacíos dispersos. Esta representación define un espacio de  $4^{20}$  puntos, mejorando la resolución en comparación con el CGR tradicional, que puede representar hasta aproximadamente 15 dígitos. Aunque la BGR no ocupa el mismo espacio que el CGR, aún preserva una estructura geométrica autosimilar. Esto sugiere que la organización general está gobernada en gran medida por la densidad o distribución de puntos. Notablemente, en el espacio CGR, los puntos no se superponen, y por lo tanto aplicar opacidad no es estrictamente necesario para resaltar regiones densas.



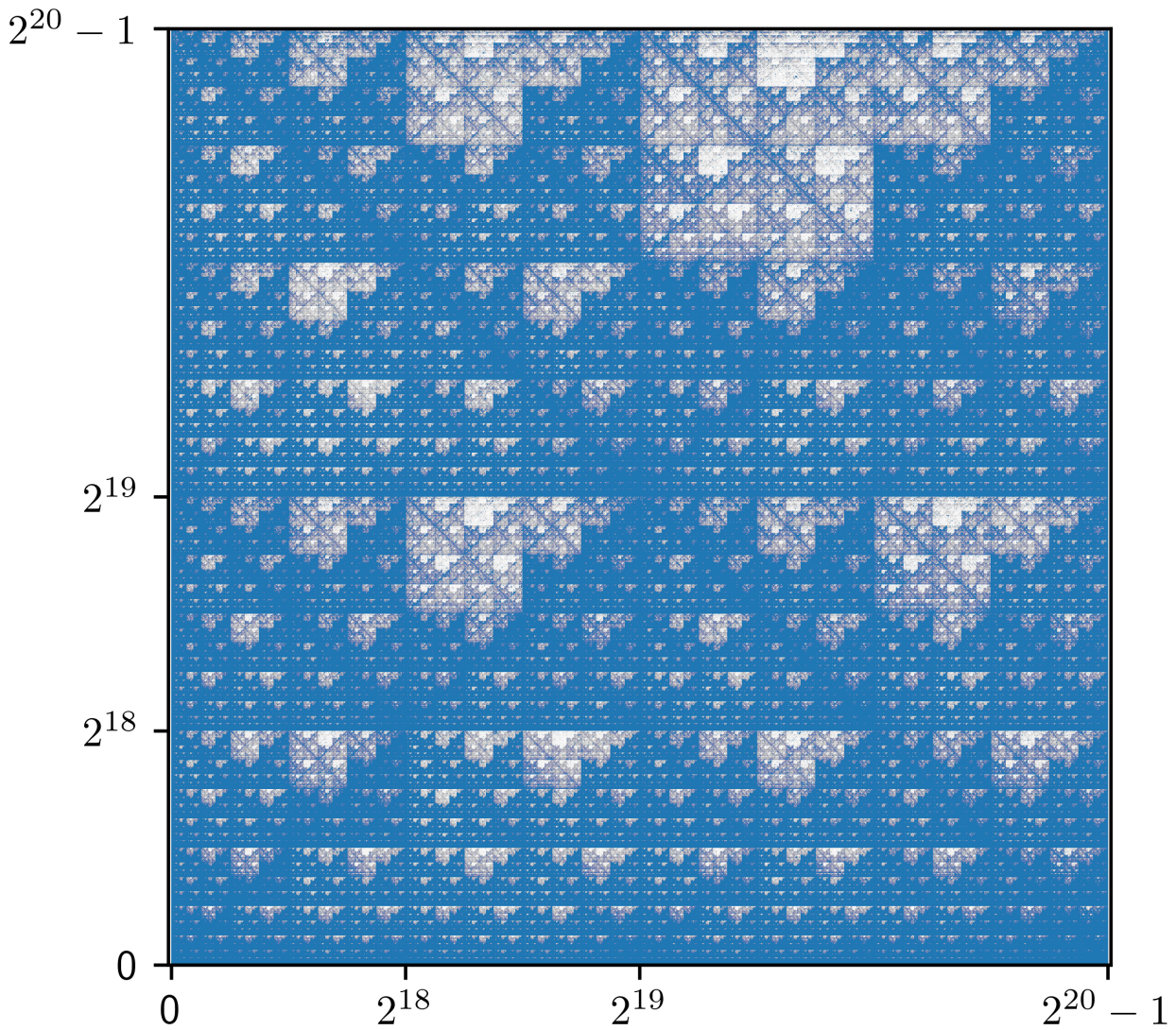
**Figura 4.4:** Cada diagrama muestra los valores posibles de  $x_i$  en el CGR, donde  $x_i$  representa la posición del  $i$ -ésimo punto, y sus ubicaciones correspondientes en la BGR para  $M = i$ . Esta correspondencia permite interpretar la BGR como una matriz de frecuencia que cuenta el número de visitas a cada punto, donde cada índice de matriz corresponde a una ubicación espacial en el plano.

## Variaciones y Rotaciones

El modelo establece la longitud de partición  $M$  para dividir secuencias en subsecuencias. En los CGR, la lectura secuencial de la SG requiere considerar variaciones en subsecuencias y sus rotaciones cíclicas cuando ocurren secuencias repetitivas.

Las variaciones (polimorfismos de un solo nucleótido) ocurren cuando secuencias de longitud idéntica difieren en un nucleótido (aunque pueden existir múltiples diferencias, este trabajo se enfoca en cambios únicos). Estas alteran la posición en el CGR dependiendo de la ubicación de la variación. Las rotaciones cíclicas corresponden a desplazamientos de nucleótidos (hacia adelante o hacia atrás) en la secuencia. Por ejemplo, las rotaciones de ACG son CGA y GAC. Cualquier secuencia de precisión  $M$  que regresa a su punto de partida después de  $M$  pasos se define como cíclica.

En el espacio BGR de tamaño  $M$ , las secuencias cíclicas ocupan  $M$  de los  $4^M$  valores posibles. Las



**Figura 4.5:** BGR con  $M = 20$  para la secuencia genómica humana completa. Los puntos se han atenuado para resaltar mejor el patrón autosimilar, como se hizo en la representación CGR mostrada en la Figura 4.1. Esta representación existe en el espacio BGR, y por lo tanto no exhibe el sesgo diagonal típico del CGR. Aunque los espacios CGR y BGR son diferentes, la estructura autosimilar se preserva en ambos.

secuencias repetitivas en tándem en genomas raramente son perfectas, ya que exhiben polimorfismos. Las variaciones generan desviaciones transitorias en trayectorias cíclicas, pero estas se restauran posteriormente. Los puntos de mayor frecuencia definen la secuencia cíclica subyacente. Si la longitud real de la secuencia cíclica excede  $M$ , las frecuencias se distribuyen uniformemente, necesitando aumentar  $M$  hasta capturar la secuencia cíclica máxima. Alternativamente, los valores grandes de  $M$  pueden inicializarse y reducirse hasta lograr la resolución óptima.

En representaciones BGR de tamaño  $M$  (derivadas de la partición de la secuencia original), las secuencias repetitivas permanecen fijas en puntos específicos, desviándose sólo cuando ocurren variaciones. Esta distinción es relevante porque:

1. Los genomas de la misma especie difieren a través de variaciones menores relativas a la longitud

total.

2. Las secuencias repetitivas exhiben menor complejidad fractal/multifractal al ocupar regiones reducidas en la representación, haciendo que otras secciones genómicas sean más relevantes en este contexto.

### Lagunas

Las subsecuencias cortas ( $\geq 8-9$  nucleótidos) ausentes en genomas humanos generan lagunas en los CGRs. Al recorrer la secuencia nucleótido por nucleótido, se observan  $4^n - 1$  celdas no ocupadas para resolución  $n$ . Por el contrario, en BGR de tamaño  $M$ , una laguna aparece solo si ninguna subsecuencia de longitud  $M$  coincide con esa posición. Esto implica que secuencias más largas que contienen la subsecuencia ausente tampoco pueden existir, generando patrones de lagunas autosimilares.

Identificar estas secuencias ausentes permite filtrar secuencias de codones, que se traducen a aminoácidos formando péptidos o polipéptidos durante la síntesis de proteínas a través del ARNm.

## 4.3. Codificación de la información de la SG

En esta sección, codificamos una SG usando la BGR siguiendo el esquema de Yin. La SG se divide en bloques  $B_\alpha$ , donde  $\alpha = 0, 1, \dots, p(M)$ . Cada bloque consiste en  $M$  nucleótidos, excepto el último, que contiene  $q(M) < M$  nucleótidos. Aquí,  $M$  es un parámetro fijo elegido tal que:

$$\begin{aligned} N + 1 &= M \cdot [p(M) - 1] + q(M) \\ p(M) &= \left\lfloor \frac{N + 1}{M} \right\rfloor \\ q(M) &= (N + 1) \bmod M \end{aligned} \tag{4.25}$$

La SG extendida, incluyendo el pseudo-nucleótido  $V_0$ , se segmenta como sigue:

$$\begin{aligned} \text{SG Extendida} &= (\vec{V}'_0, \vec{V}'_1, \dots, \vec{V}'_{M-1}) \\ &\quad (\vec{V}'_M, \vec{V}'_{M+1}, \dots, \vec{V}'_{2M-1}) \\ &\quad \dots (\vec{V}'_{pM}, \vec{V}'_{pM+1}, \dots, \vec{V}'_N) \end{aligned} \tag{4.26}$$

Usando el esquema de Yin, cada bloque  $B_\alpha$  se codifica con un vector entero  $\vec{N}_\alpha$ :

$$\begin{aligned} \vec{N}_\alpha &= \sum_{k=0}^{M-1} 2^k \vec{V}'_{\alpha M+k}, \quad \alpha = 0, \dots, p-1 \\ \vec{N}_p &= \sum_{k=0}^q 2^k \vec{V}'_{pM+k} \end{aligned} \tag{4.27}$$

El vector asociado con toda la SG es:

$$\begin{aligned}\vec{Y}_N &= \sum_{j=0}^N 2^j \vec{V}_j' \\ &= \sum_{\alpha=0}^{p-1} 2^{\alpha M} \vec{N}_\alpha + 2^{pM} \vec{N}_p\end{aligned}\tag{4.28}$$

Así, el caminante digital de grano grueso se describe por los  $p + 1$  vectores enteros:

$$\begin{aligned}&(\vec{N}_0, \vec{N}_1, \dots, \vec{N}_{p-1}, \vec{N}_p) \\ &= ((N_0^x, N_0^y), (N_1^x, N_1^y), \dots, (N_p^x, N_p^y))\end{aligned}\tag{4.29}$$

Adicionalmente, un vector  $\vec{A} = (N + 1, M)$  almacena el número total de nucleótidos y el tamaño de bloque:

$$\begin{aligned}\vec{A} &= (N + 1, M) \\ p(M) &= \left\lfloor \frac{N + 1}{M} \right\rfloor \\ q(M) &= (N + 1) \text{ mód } M\end{aligned}\tag{4.30}$$

Para codificar una subsecuencia de los primeros  $R$  nucleótidos:

$$\begin{aligned}R &= M\alpha_0 + s - 1 \\ \alpha_0 &= \left\lfloor \frac{R}{M} \right\rfloor \\ s &= R \text{ mód } M\end{aligned}\tag{4.31}$$

El vector correspondiente  $\vec{Y}_R$  es:

$$\begin{aligned}\vec{Y}_R &= \sum_{j=0}^R 2^j \vec{V}_j' \\ &= \sum_{\alpha=0}^{\alpha_0-1} 2^{\alpha M} \vec{N}_\alpha + 2^{\alpha_0 M} \vec{N}_s\end{aligned}\tag{4.32}$$

donde el bloque parcial está codificado por:

$$\vec{N}_s = \sum_{k=0}^{s-1} 2^k \vec{V}_{k+\alpha_0 M}'\tag{4.33}$$

Por lo tanto, la subsecuencia está codificada por el conjunto:

$$(N_0, N_1, \dots, N_s), \quad N_s = \sum_{k=0}^s 2^k V_{\alpha_0 M + k}\tag{4.34}$$

Esto requiere solo  $\alpha_0 + 1$  vectores enteros.

Considere codificar una secuencia extendida de 30 nucleótidos más un nucleótido cero:

$$\begin{array}{c}
 V_0 A C T G G T A \\
 T T G T A C T C \\
 C T A A C C C T \\
 A G G T T G C, \quad (3+1)
 \end{array} \tag{4.35}$$

El vector inicial es:

$$I = (31, 8)^T \tag{4.36}$$

indicando 31 nucleótidos divididos como:

$$\begin{aligned}
 p &= \left\lfloor \frac{31}{8} \right\rfloor = 3 \\
 q &= 31 \bmod 8 = 7 \\
 \alpha &= (0, 1, 2, 3), \quad (3+1)
 \end{aligned} \tag{4.37}$$

Cada grupo se codifica usando conversión binaria-a-decimal. Por ejemplo:

$$\begin{aligned}
 N_{0,x} &= 01111001_2 = 121_{10} \\
 N_{0,y} &= 00110101_2 = 53_{10} \\
 \vec{N}_0 &= (121, 53)
 \end{aligned} \tag{4.38}$$

Así, la codificación completa es:

$$\begin{aligned}
 I &= (31, 8) \\
 \vec{N}_0 &= (121, 53), \quad \vec{N}_1 = (79, 164) \\
 \vec{N}_2 &= (130, 113), \quad \vec{N}_3 = (62, 102) \\
 &\Rightarrow (3+1)
 \end{aligned} \tag{4.39}$$

Por lo tanto, toda la secuencia de 31 nucleótidos se codifica usando solo 4 vectores enteros de “grano grueso”, como se muestra en la Tabla 4.1.

**Tabla 4.1:** Segmentación de la secuencia (4.35), junto con sus codificaciones binarias correspondientes y representaciones decimales para cada bloque  $\alpha$ .

|                      | $\alpha = 0$                                         | $\alpha = 1$                                         | $\alpha = 2$                                         | $\alpha = 3$                                         |
|----------------------|------------------------------------------------------|------------------------------------------------------|------------------------------------------------------|------------------------------------------------------|
| SGE                  | $V_0$ ACTGGTA                                        | TTGTACTC                                             | CTAACCCT                                             | AGGTTGC                                              |
| $\vec{P}_\alpha$     | $\begin{pmatrix} 10011110 \\ 10101100 \end{pmatrix}$ | $\begin{pmatrix} 11110010 \\ 00100101 \end{pmatrix}$ | $\begin{pmatrix} 01000001 \\ 10001110 \end{pmatrix}$ | $\begin{pmatrix} 01111110 \\ 01100111 \end{pmatrix}$ |
| $\vec{P}_{N_\alpha}$ | $\begin{pmatrix} 121 \\ 53 \end{pmatrix}$            | $\begin{pmatrix} 79 \\ 164 \end{pmatrix}$            | $\begin{pmatrix} 130 \\ 113 \end{pmatrix}$           | $\begin{pmatrix} 62 \\ 102 \end{pmatrix}$            |

Para la decodificación, el proceso se invierte leyendo la Tabla 4.1 de abajo hacia arriba. Comenzando desde los vectores  $\{\vec{N}_\alpha\}$ , cada uno se convierte a binario (los primeros tres con 8 dígitos significativos, el último con 7). Luego, los nucleótidos para cada bloque se identifican usando la ecuación correspondiente, y concatenándolos, la secuencia original (4.35) se reconstruye sin pérdida de información.

Usando (4.27) (o su contraparte gráfica), los vectores  $\{\vec{N}_\alpha\}$  determinan los puntos gráficos  $\vec{Y}_j$  como sigue. Dados  $\vec{N}_0, \vec{N}_1, \vec{N}_2$  y  $\vec{N}_3$ , se obtiene

$$\begin{aligned}\vec{Y}_7 &= \vec{N}_0; \\ \vec{Y}_{15} &= \vec{N}_0 + 2^8 \vec{N}_1, \\ \vec{Y}_{23} &= \vec{N}_0 + 2^8 \vec{N}_1 + 2^{16} \vec{N}_2, \\ \vec{Y}_{30} &= \vec{N}_0 + 2^8 \vec{N}_1 + 2^{16} \vec{N}_2 + 2^{24} \vec{N}_3.\end{aligned}$$

En el esquema gráfico de Yin, cada punto codifica solo su propio bloque:

$$\vec{P}_7 = \vec{N}_0, \quad \vec{P}_{15} = 2^8 \vec{N}_1, \quad \vec{P}_{23} = 2^{16} \vec{N}_2, \quad \vec{P}_{30} = 2^{24} \vec{N}_3.$$

En la Figura 4.6 se utiliza como ejemplo la RGB con  $M = 1$  y  $M = 2$  aplicada al cromosoma 1 humano. La figura ilustra cómo se representan los puntos obtenidos mediante la RGB, así como la subsecuencia correspondiente asociada a cada punto. Para cada valor de  $M$  existen  $4^M$  posibles secuencias, las cuales se ubican en la matriz siguiendo la asignación binaria y luego convertidas a base 10. De este modo, a cada subsecuencia se le asocia un único punto en el plano, que además corresponde a una coordenada entera positiva  $(i, j)$  en la matriz, donde  $i$  indica la columna y  $j$  la fila.

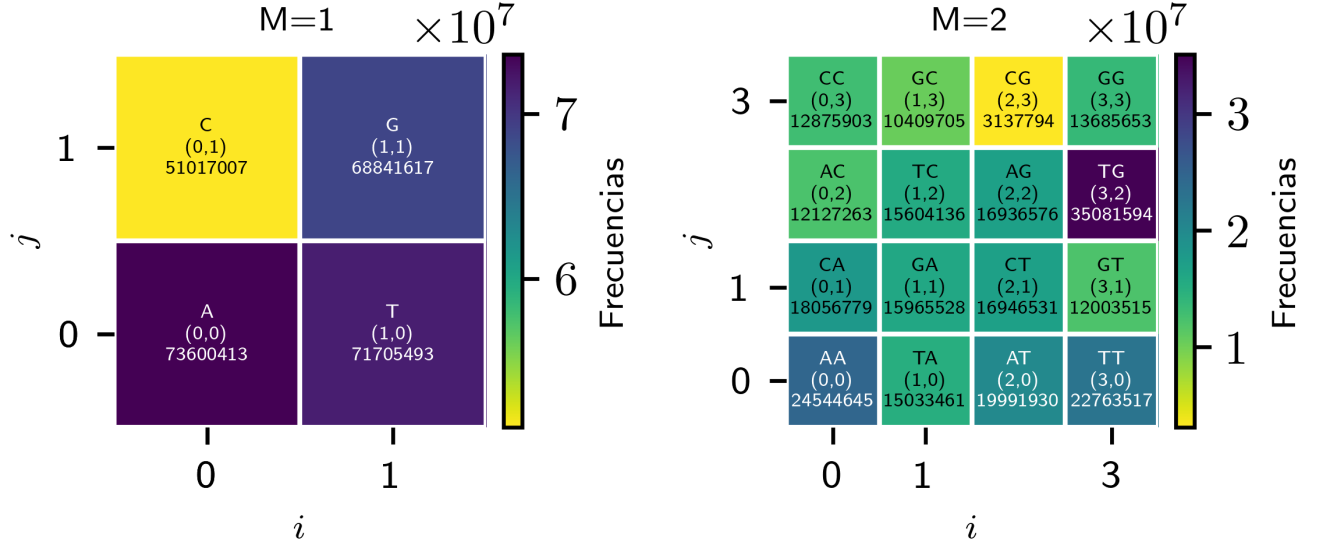
La organización definida por la RGB puede entenderse como una partición del plano en regiones determinadas por los puntos centrales asociados a las subsecuencias. A cada una de estas regiones se le asigna el número de veces que aparece su subsecuencia correspondiente en la SG, lo que permite construir una matriz de frecuencias. Estas frecuencias se visualizan mediante colores, diferenciando las regiones de acuerdo con la cantidad de apariciones.

Como se observa, la organización de las secuencias sigue la jerarquía que se lee de derecha a izquierda. Por ejemplo, todas las secuencias que terminan en la base A pertenecen a una misma región, la cual se subdivide en cuatro subregiones correspondientes a las bases precedentes AA, CA, GA y TA, manteniendo el orden definido. Como cada color en la gráfica representa la frecuencia de aparición en la SG, se genera así una distribución bidimensional de las SGs.

## 4.4. Datos y Metodología

Para ilustrar el funcionamiento del método BGR, utilizamos como ejemplo la distribución obtenida a través del CGR para subsecuencias de longitud  $M$ . Para ello, se aplica la BGR  $M$  veces usando





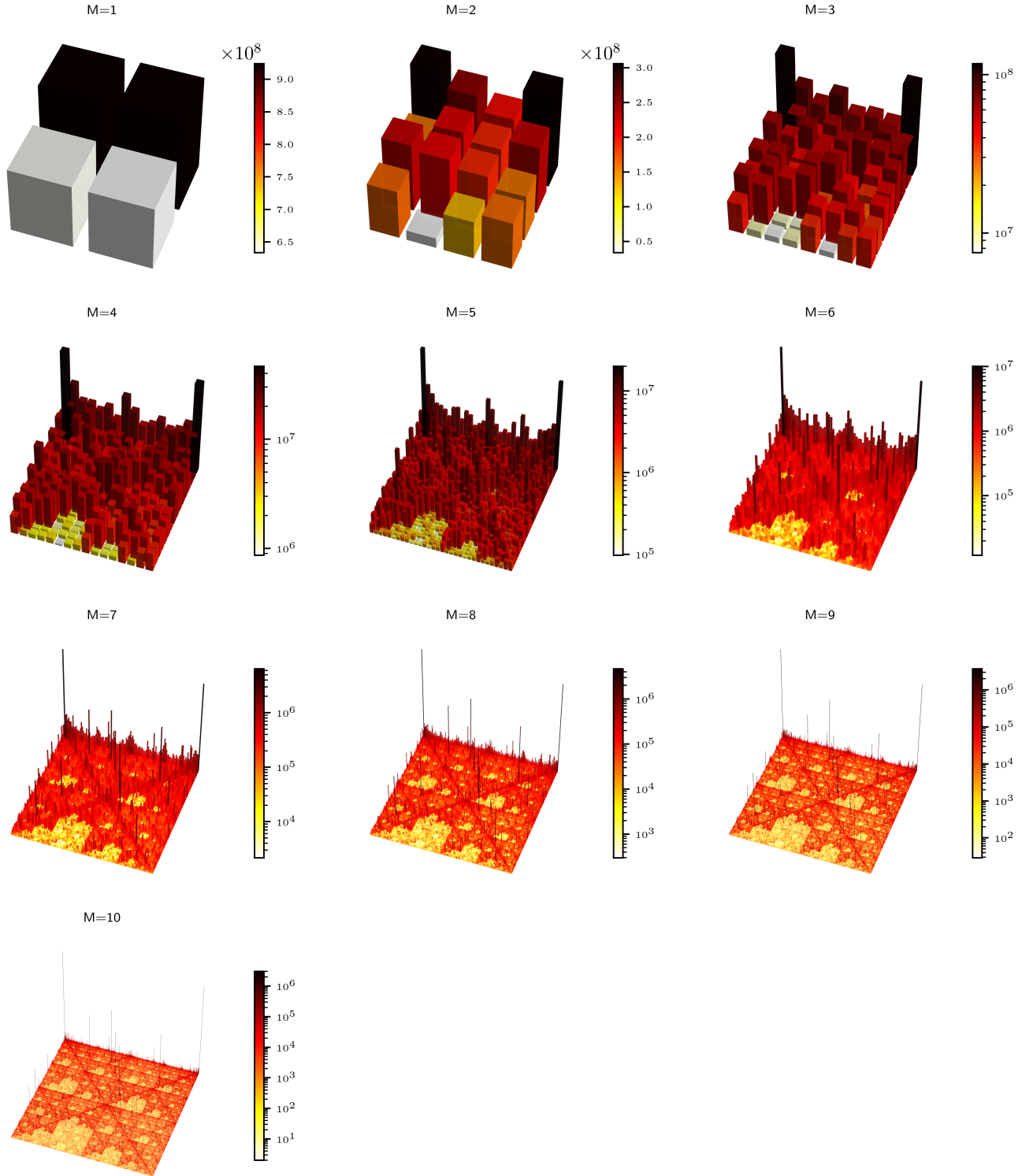
**Figura 4.6:** Distribuciones obtenidas mediante la RGB con  $M = 1$  y  $M = 2$  para el cromosoma 1 humano. En este esquema se ilustra cómo la RGB mapea las secuencias de longitud  $M$  en el plano. Cada subsecuencia se representa mediante un punto  $(i, j)$  en el dominio  $[0, 2^M - 1]^2$ , donde  $i$  y  $j$  son enteros positivos. La matriz resultante está compuesta por los conteos de ocurrencia de cada subsecuencia dentro de la SG. En el esquema, cada región está coloreada según su frecuencia, mostrando también el número de apariciones correspondiente.

ventanas deslizantes que recorren toda la secuencia genómica, permitiéndonos contar los puntos asociados con cada subsecuencia. Este procedimiento es equivalente a calcular la frecuencia de todas las subsecuencias posibles de longitud  $M$  dentro del genoma humano completo *T2T-CHM13v2.0* [8]. Tanto CGR como BGR producen  $L - (M - 1)$  puntos, correspondientes a cada subsecuencia de longitud  $M$ .

La matriz de frecuencias obtenida mediante BGR también puede utilizarse para derivar las distribuciones correspondientes a subsecuencias más cortas, con longitud  $M' < M$ . Para hacerlo, la matriz se subdivide de acuerdo con el nuevo valor  $M'$ , como se ilustra en la Figura 4.2, y se suman las frecuencias dentro de cada subregión, obteniendo así la distribución desde la posición  $M$  hasta el final de la secuencia. Luego se aplica la BGR a subsecuencias de longitud  $M'$  desde la primera posición hasta la posición  $M$ , y las dos matrices resultantes se suman, reconstruyendo así la distribución completa de subsecuencias de longitud  $M'$  a lo largo de toda la SG.

## 4.5. Resultados y Análisis

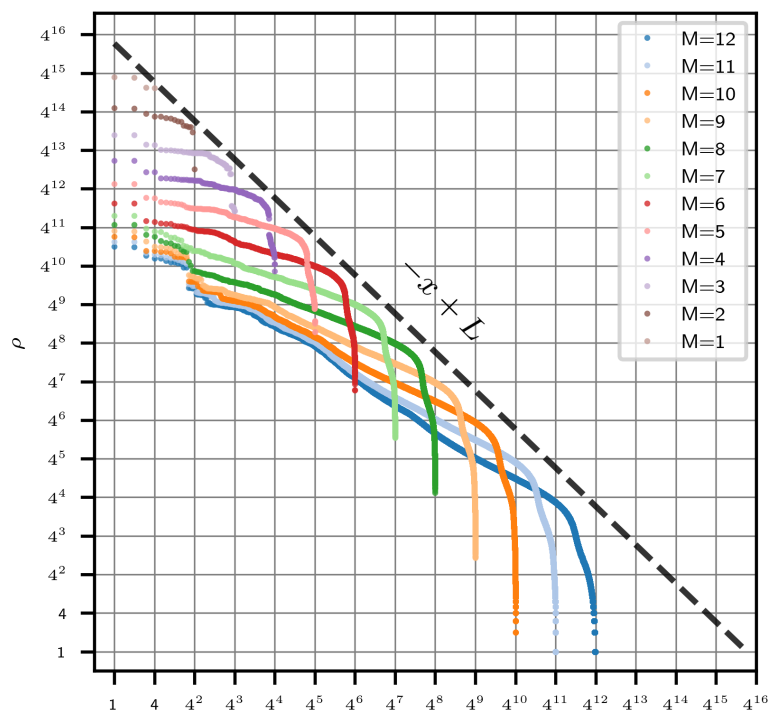
La Figura 4.7 muestra las distribuciones obtenidas para los primeros diez valores de  $M$  usando el método BGR aplicado a la SG completa [8], que consta de 3,117,292,070 bases, incluyendo secuencias cromosómicas y ADN mitocondrial. La distribución general representa claramente la superposición de las distribuciones separadas, donde cada cromosoma y el ADN mitocondrial tienen representaciones y distribuciones distintas.



**Figura 4.7:** Distribuciones BGR para los primeros diez valores de  $M$  en la SG humana completa *T2T-CHM13v2.0*. Cada distribución se obtiene contando puntos en una malla de  $2^M \times 2^M$  (Figura 4.5). Para resaltar las variaciones a gran escala en los resultados, se empleó una escala logarítmica, que también ayuda a visualizar el patrón autosimilar en la BGR. Los picos más altos corresponden a la base T en la esquina superior izquierda y la base A en la esquina superior derecha de cada representación, siguiendo el esquema de asignación de bases en la Figura 4.1.

Diferentes conjuntos de puntos en la BGR se caracterizan por su frecuencia a lo largo de la SG. Los conjuntos de alta frecuencia son pequeños, consistiendo principalmente en un solo punto, lo que no define un comportamiento autosimilar. Sin embargo, dentro de un amplio rango de escalas (Figura 4.8) gobernado por leyes de potencia, estos corresponden a un extremo en el espectro multifractal

Los resultados para el ensamblaje cromosómico completo muestran que la SG contiene grandes regiones repetitivas, compuestas principalmente por bases A o T, predominantes en la mayoría de los cromosomas. Muchas secuencias asociadas con puntos de alta frecuencia son polimorfismos que difieren en un solo nucleótido y se ubican en las mismas secciones de la SG. Por lo tanto, los segmentos altamente repetitivos pueden formarse por repeticiones mixtas de diferentes polimorfismos.

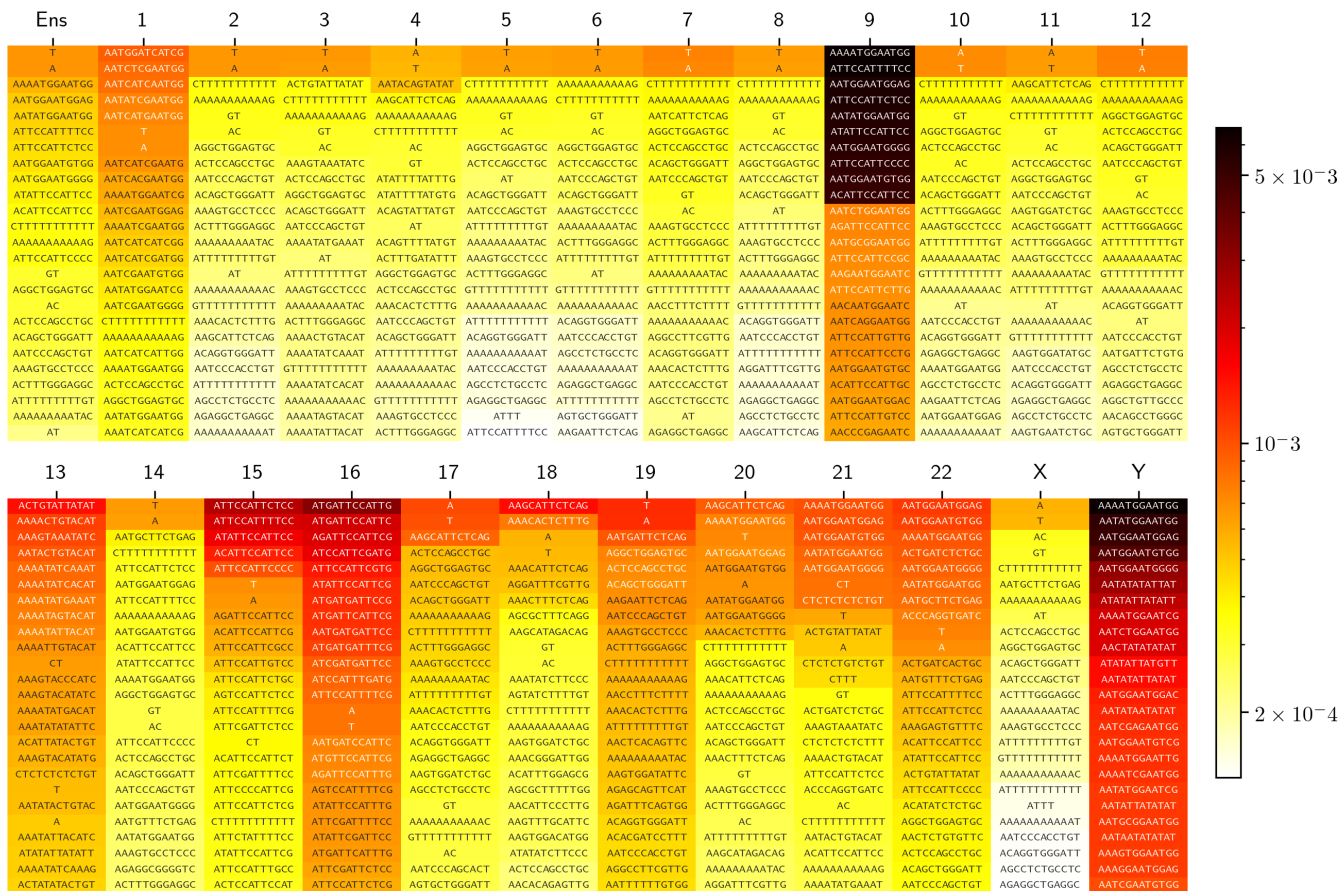


**Figura 4.8:** BGR ordenada por densidad para los primeros valores de  $M$ . Para cada valor de  $M$ , la distribución correspondiente se ordena en orden descendente y los resultados se muestran en escala log-log. Esta representación revela un amplio rango de variación de escala en la decadencia de frecuencia, destacando la presencia de fuerte heterogeneidad en las distribuciones.

La secuenciación puede realizarse para identificar las ubicaciones de cada subsecuencia de longitud  $M$  a lo largo de la SG, cromosoma por cromosoma. En la Figura 4.9, se listan las secuencias de mayor densidad en relación con sus respectivos conteos de bases, y la Figura 4.10 muestra su localización a lo largo de cada cromosoma junto con las frecuencias acumuladas para  $M = 12$ . Notablemente, cada cromosoma exhibe patrones de alta frecuencia distintos, variando significativamente tanto las secuencias específicas como sus ubicaciones genómicas. Los valores de alta frecuencia corresponden al subconjunto más pequeño de toda la distribución de frecuencia, lo que significa que los picos más altos típicamente representan solo una subsecuencia de longitud  $M$ .

Las rotaciones cíclicas también se consideraron en las Figuras 4.9 y 4.10, mostrando que los picos

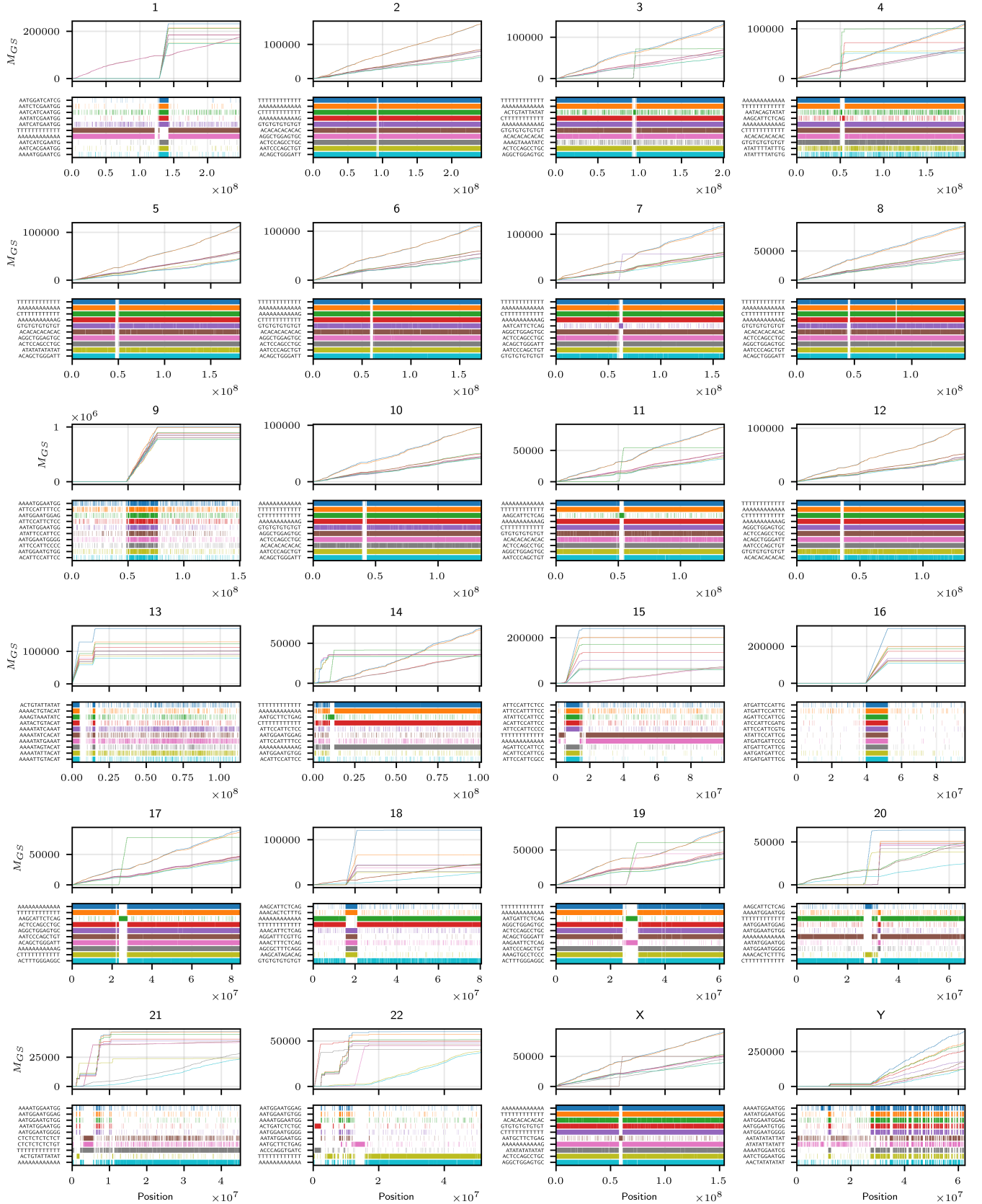
#### 4.5. RESULTADOS Y ANÁLISIS



**Figura 4.9:** Secuencias de mayor frecuencia en la SG. Estas secuencias corresponden a los valores más altos en la distribución BGR para  $M = 12$  calculada a partir del ensamblaje *T2T-CHM13v2.0*. El conjunto de picos en la distribución está compuesto principalmente por secuencias únicas de longitud  $M$ . Cuando una secuencia consiste en repeticiones de un motivo más corto, solo se reporta su unidad repetitiva mínima.

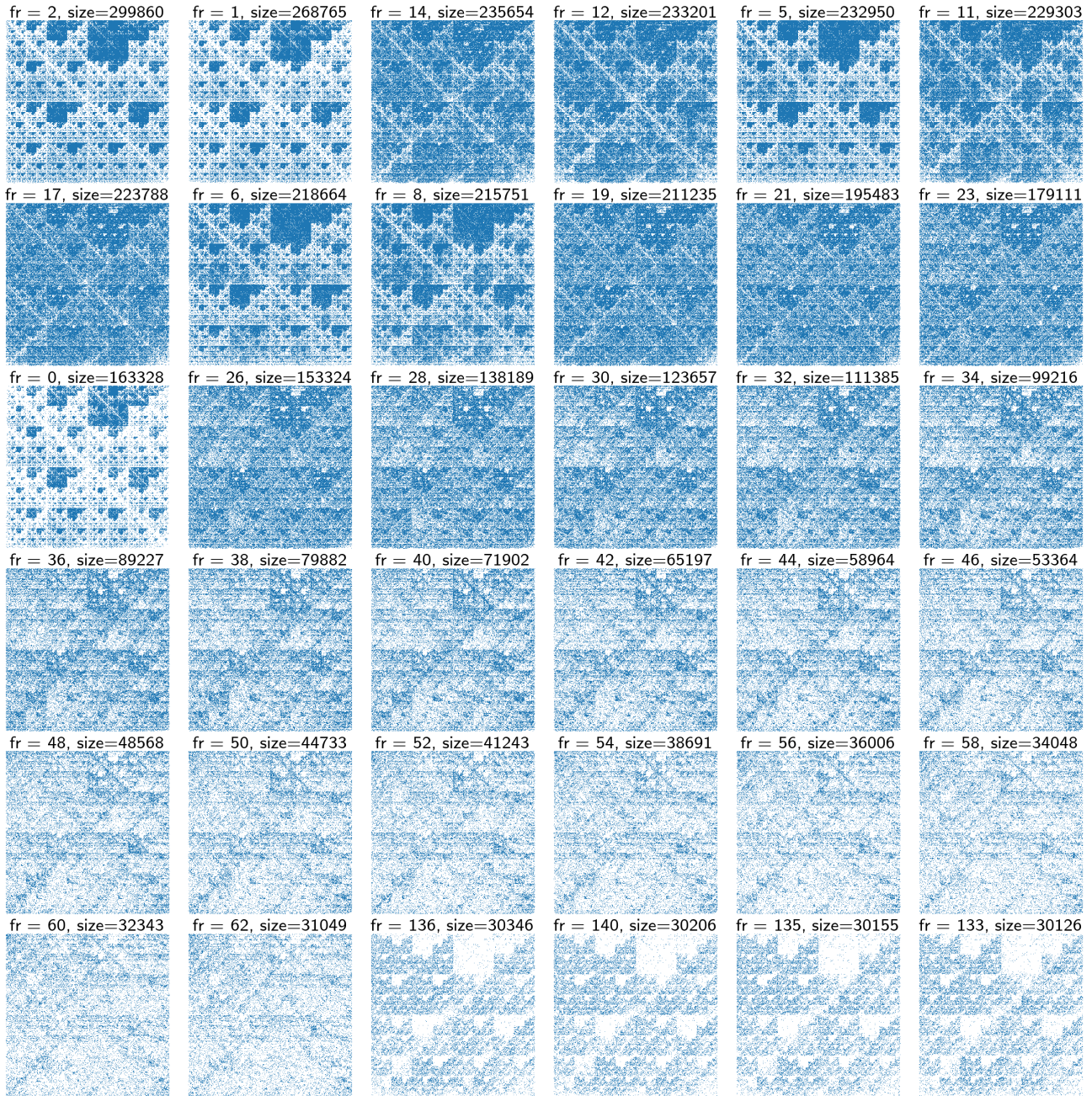
no se alinean perfectamente con variantes rotacionales. Esto indica que es poco probable que existan repeticiones perfectas largas de una sola secuencia en regiones repetitivas. En cambio, estas regiones representan zonas altamente polimórficas de la SG, compuestas por subsecuencias imperfectas pero recurrentes regularmente, como se discutió previamente.

Las distribuciones acumuladas revelan un crecimiento aproximadamente lineal en regiones con predominio de secuencias de alta frecuencia en diferentes cromosomas, consistente con la presencia de regiones genómicas repetitivas. En ciertos cromosomas—específicamente 1, 9, 13, 15, 16, 18, 20, 21, 22 e Y—hay un cambio notable hacia secuencias distintas de aquellas compuestas predominantemente por A's y T's. Estas incluyen altos contenidos de motivos como ATTCC/GGAAT o TTAGGG, que a menudo se asocian con clases y subclases de ADN satélite. Este ADN, que constituye heterocromatina en los centrómeros y otras regiones, consiste principalmente en arreglos largos de secuencias repetidas en tandem de alto número de copias. Grandes trectos de heterocromatina típicamente muestran una organización en mosaico de diferentes familias de ADN satélite, ocasionalmente interrumpidas por elementos transposones. La estructura del ADN satélite varía, con unidades repetitivas que van desde secuencias muy cortas (menos de 10 nucleótidos) hasta motivos moderadamente complejos que se



**Figura 4.10:** Frecuencia alta (pico BGR) y secuenciación acumulativa  $M_{GS}$  para  $M = 12$  a través de los cromosomas. Para cada cromosoma, las posiciones de puntos de alta frecuencia se localizan a lo largo de toda su SG. Cada secuencia de longitud doce también representa todas sus rotaciones cíclicas. Los cromosomas se diferencian principalmente por la distribución de regiones altamente repetitivas a lo largo de sus secuencias, aunque pueden observarse similitudes parciales entre determinados cromosomas.





**Figura 4.11:** Se muestra una muestra de conjuntos en la BGR con frecuencias idénticas para  $M = 12$  en la SG humana completa. Cada conjunto se define por el número de veces que aparece en la BGR y por el número de puntos con tales frecuencias. En general, los conjuntos con frecuencias más bajas tienden a ser más grandes en tamaño, especialmente cuando se considera cada cromosoma independientemente (Figura 4.12). Estos conjuntos exhiben diferentes patrones espaciales asociados con la SG, algunos más predominantes que otros, lo que se refleja en sus distribuciones correspondientes.

extienden más de 100 nucleótidos de longitud.

En frecuencias bajas, los conjuntos de puntos son relativamente grandes—del orden de  $10^5$  para los más grandes—donde ya puede identificarse autosimilitud (Figura 4.11). Aunque los patrones pueden parecer similares, no son idénticos y corresponden a secuencias distintas. Cada patrón cubre el espacio de manera diferente, algunos exhibiendo distribuciones más homogéneas dependiendo de su frecuencia y tamaño. A medida que el tamaño del conjunto disminuye y la frecuencia aumenta, la presencia del dinucleótido CG se reduce. En consecuencia, la tendencia en la SG es hacia una cobertura espacial reducida a medida que aumenta la frecuencia.

En la mayoría de los cromosomas (Figura 4.12), los conjuntos de baja frecuencia se dispersan dentro de regiones ricas en CG. Sin embargo, para algunos cromosomas, las distribuciones son más homogéneas, dificultando identificar visualmente tendencias relacionadas con la composición de nucleótidos. Siguiendo este comportamiento, es probable que, a medida que aumenta la frecuencia, los conjuntos tiendan a agruparse en posiciones más cercanas, similar a lo observado en las frecuencias más altas.

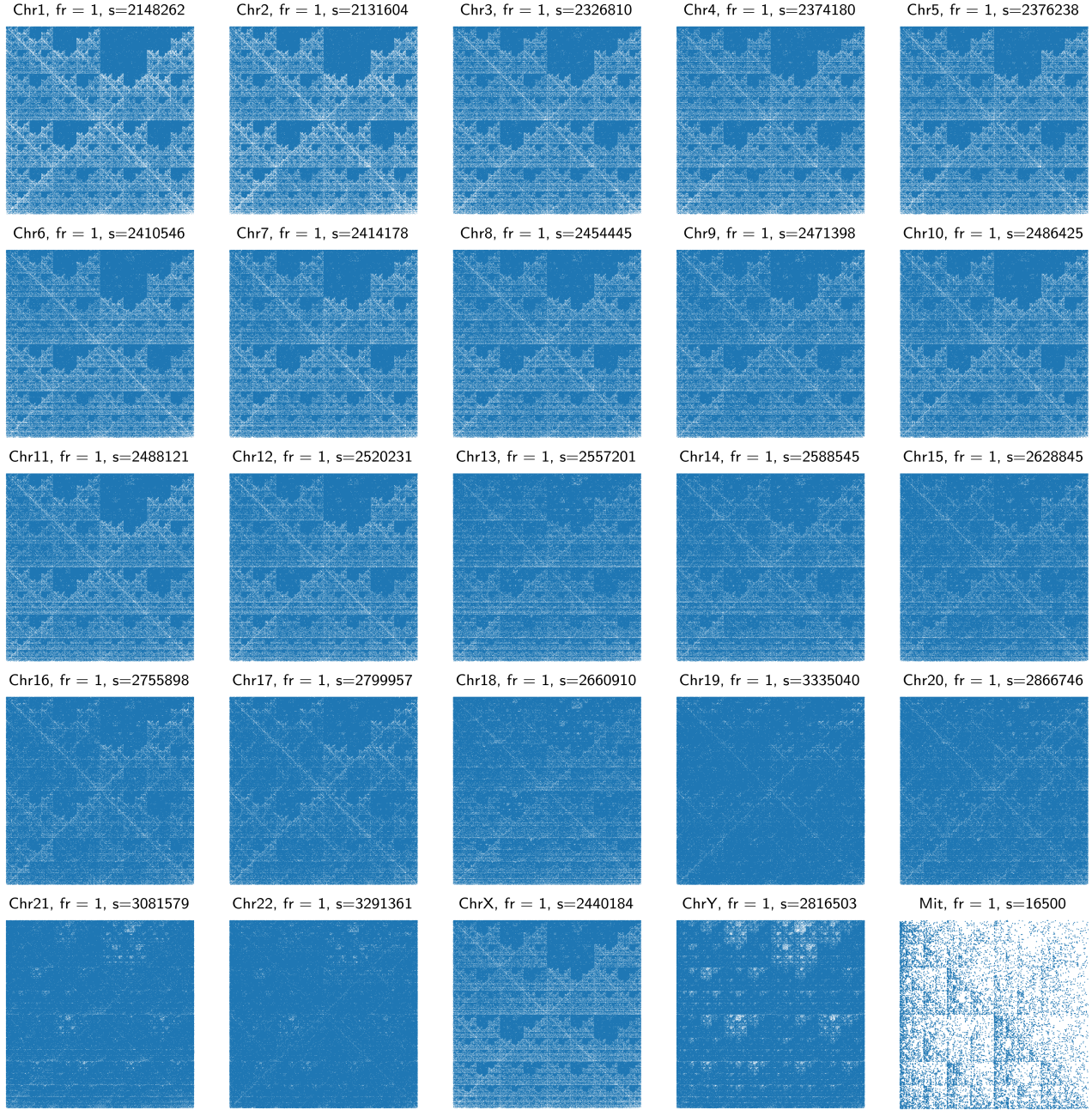
Esta característica también es evidente cuando las secuencias se separan por cromosomas. La Figura 4.12 muestra el conjunto más grande—por número de puntos con la misma frecuencia—para cada cromosoma y ADN mitocondrial. En todos los casos, este conjunto corresponde a frecuencia 1 para  $M = 12$ . Notablemente, las regiones vacías, que representan secuencias ausentes de la SG, forman conjuntos incluso más grandes que aquellos con frecuencia 1 en todos los cromosomas.

Cuando se ordenan por tamaño de conjunto, observamos que el tamaño aumenta con la frecuencia. En la Figura 4.13, realizamos una secuenciación de las posiciones de secuencias correspondientes a cada punto en los conjuntos más grandes para cada cromosoma, comenzando con el más grande y saltando cada 14 conjuntos por tamaño. Una de las primeras características que podemos caracterizar—consistente con las expectativas—es que, en cada cromosoma, la densidad de estas secuencias disminuye o se desvanece en regiones altamente repetitivas. Fuera de estas regiones, hay una clara tendencia a que las secuencias se distribuyan linealmente a través del genoma, por lo que la contribución de cada conjunto puede ser significativamente mayor a medida que aumenta la frecuencia, aunque existe cierta variación entre cromosomas.

También destacamos que el ensamblaje completo contiene un número mucho mayor de conjuntos que lo observado cuando se analiza por cromosoma, y que el ADN mitocondrial contiene solo dos conjuntos distintos. Para  $M = 12$ , el ensamblaje completo tiene 17,174 conjuntos diferentes, mientras que el número de conjuntos distintos por cromosoma es el siguiente: 4838, 3636, 3323, 3219, 3166, 2986, 3332, 2663, 4086, 2968, 2867, 3038, 2822, 2690, 3250, 3731, 3017, 2511, 2930, 2311, 1731, 2624, 2865 y 3458.

Aunque parece haber cierta dependencia del número de conjuntos con la longitud de la secuencia, esta dependencia no es del todo clara. Por ejemplo, el cromosoma Y tiene más conjuntos que el



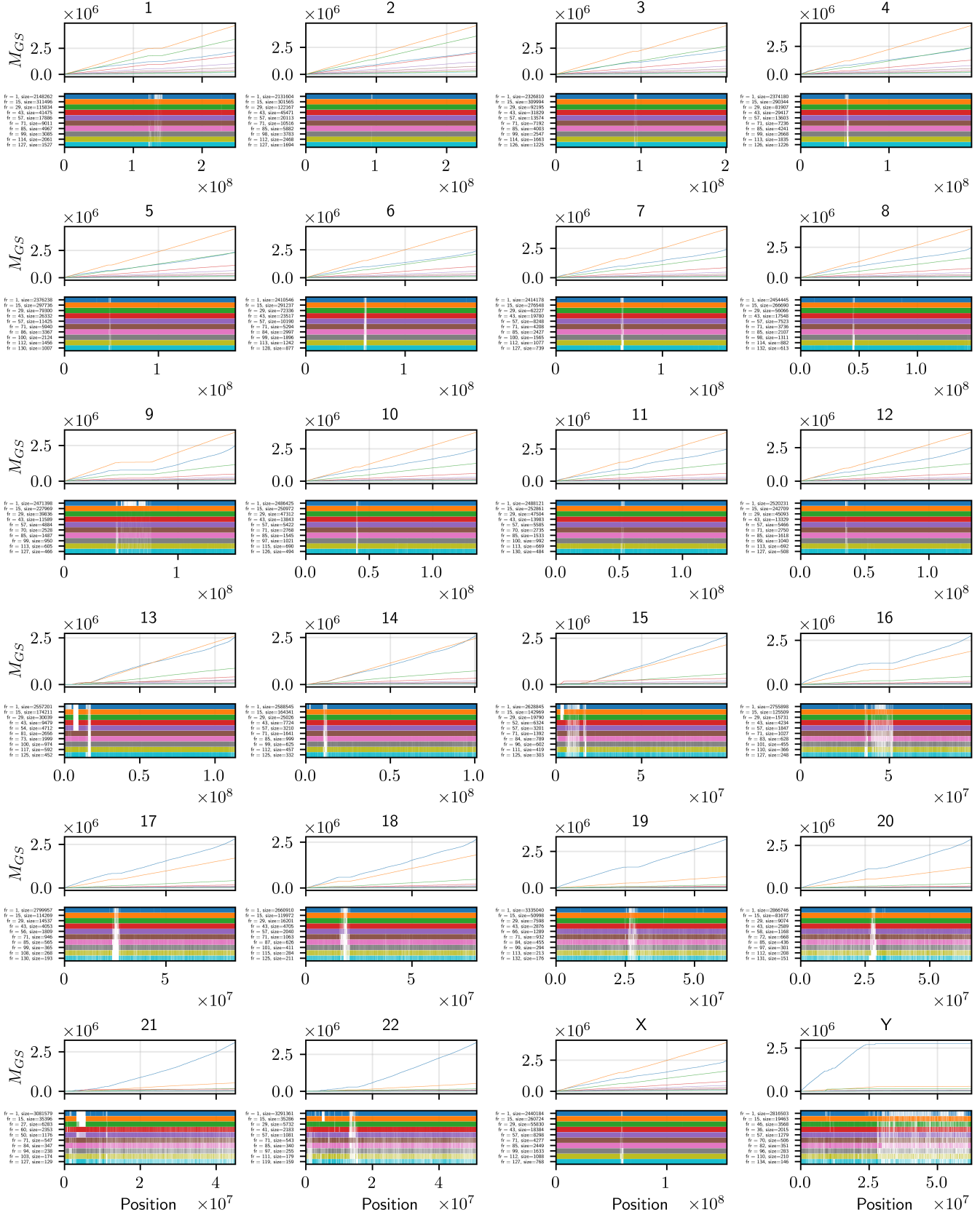


**Figura 4.12:** Los conjuntos más grandes de frecuencias idénticas para cada cromosoma y ADN mitocondrial corresponden a frecuencia 1 a través de todas las SGs. Estos conjuntos exhiben patrones similares, predominantemente ubicados en regiones ricas en CG en la mayoría de los cromosomas, con la excepción del ADN mitocondrial. Notablemente, el tamaño de estos conjuntos puede variar significativamente, excediendo un millón de puntos en algunos cromosomas.



cromosoma X, y el cromosoma 9 tiene el segundo mayor número después del cromosoma 1. Estos hallazgos indican que el tamaño de los conjuntos en los cromosomas aumenta drásticamente, alcanzando el orden de  $10^6$ , con el conjunto más grande comprendiendo 3,335,040 puntos en el cromosoma 19.

Para el ADN mitocondrial y  $M = 12$ , hay poca variabilidad en términos de conjuntos de secuencias con frecuencias idénticas, y se espera que esta tendencia se mantenga para valores mayores de  $M$ . En este caso, hay un conjunto de 16,500 puntos con frecuencia 1 y otro de 29 puntos con frecuencia 2. Dado que el ADN mitocondrial es significativamente más corto que las secuencias cromosómicas, el rango de posibilidades disminuye considerablemente a medida que  $M$  aumenta. El patrón correspondiente a secuencias que aparecen solo una vez en la SG difiere de aquellos encontrados en los esquemas previos de cromosomas o del ensamblaje completo, alineándose más estrechamente con la estructura del triángulo de Sierpinski. Esto refleja una reducción en la presencia de la base G, con una preferencia por dinucleótidos de tipo AC, correspondiendo a las regiones más densas en la Figura 4.12, que representa el ADN mitocondrial.



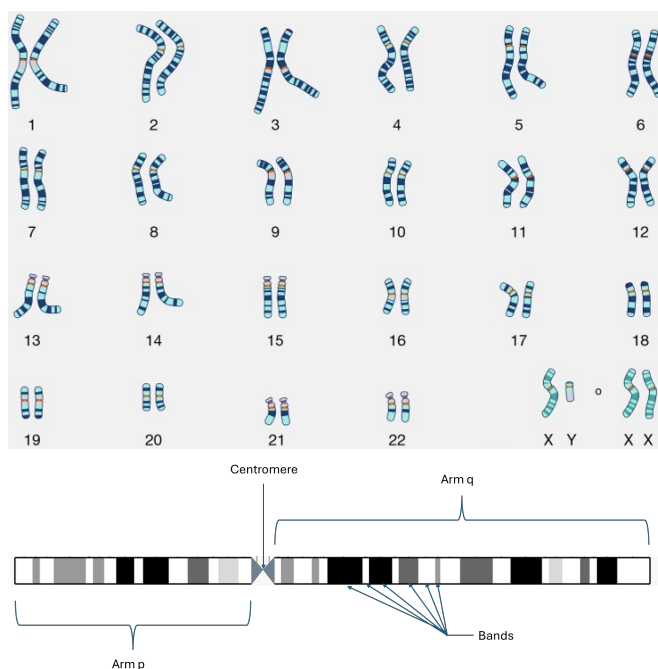
**Figura 4.13:** Secuenciación de conjuntos con frecuencias iguales clasificados por tamaño. Para cada cromosoma, se muestran las posiciones y frecuencias acumuladas a lo largo de la SG. El conjunto más grande se muestra primero, seguido por otros muestreados saltando cada 14 tamaños, ofreciendo una visión más representativa de diversos comportamientos de patrones.

## Capítulo 5

# Análisis Multifractal del Genoma Humano

### 5.0.1. Citogenética Humana

El número y la morfología de los cromosomas en una célula de una especie específica son generalmente constantes en la mayoría de las células de ese organismo (con excepción de las células reproductivas y algunas otras, como las hepáticas). Esta es una característica distintiva de cada especie. La caracterización cromosómica se realiza típicamente mediante técnicas de bandeo cromosómico, que producen bandas visibles en los cromosomas. Estas bandas crean un patrón único para cada cromosoma (ver Figura 5.1). Esta información también puede capturarse a lo largo de la secuencia genómica (SG) para cada cromosoma, y existen bases de datos especializadas donde los científicos recopilan y almacenan nuevos hallazgos basados en la SG.



**Figura 5.1:** Cariotipo de un individuo, mostrando todos los cromosomas humanos y su estructura, compuesta por cuatro partes básicas: el brazo p, el centrómero, el brazo q y las bandas a lo largo de toda la secuencia. Fuente: “Cortesía: National Human Genome Research Institute”. [20]

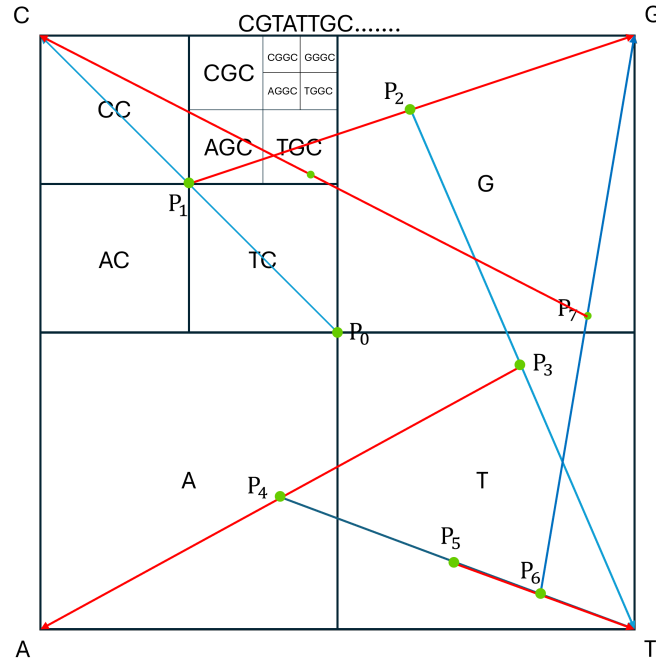
---

Como se mencionó anteriormente, las SGs humanas contienen diferentes tipos de regiones funcionales y estructurales, cada una con un papel específico en la organización, regulación y función del genoma. A continuación, se resumen brevemente las principales regiones y sus funciones, información que nos permite aproximarnos a una función para cada parte de las representaciones gráficas [19]:

- **Genes y exones:** Los genes contienen la información necesaria para producir proteínas y ARN funcionales. Los exones son las partes del gen que se transcriben y traducen en proteínas, mientras que los intrones se eliminan durante el procesamiento del ARN.
- **Promotores y elementos reguladores:** Los promotores son secuencias cercanas a los genes que regulan el inicio de la transcripción. También existen potenciadores, silenciadores y otros elementos reguladores que modulan la actividad génica, influyendo en qué genes se expresan y cuándo. Esto significa que los potenciadores aumentan la probabilidad de expresión génica, los silenciadores la reducen o previenen, y otros elementos reguladores ayudan a determinar el momento y las condiciones bajo las cuales se activan o reprimen genes específicos, permitiendo que las células respondan a diferentes necesidades y ambientes.
- **Regiones no codificantes:** Aunque estas regiones no codifican proteínas, son importantes para la regulación génica. Incluyen secuencias de ARN no codificante, como microARN y ARN largos no codificantes (lncARN), que regulan la expresión génica y otros procesos celulares.

**Regiones repetitivas:** El genoma humano contiene varias secuencias repetitivas, entre ellas:

- **Microsatélites y minisatélites:** Secuencias repetitivas cortas que pueden variar entre individuos.
- **Repeticiones intercaladas:** Incluyen elementos SINE (Elementos Nucleares Cortos Intercalados) y LINE (Elementos Nucleares Largos Intercalados), secuencias repetidas distribuidas en todo el genoma que, en algunos casos, pueden moverse a otras partes del ADN.
- **Centrómeros y télómeros:** Los centrómeros son regiones especializadas involucradas en la separación cromosómica durante la división celular. Los télómeros son secuencias repetitivas en los extremos de los cromosomas que protegen la integridad de la secuencia genómica, acortándose en cada división celular.
- **Islas CpG:** Regiones con alta densidad de dinucleótidos de citosina y guanina, comúnmente encontradas cerca de promotores génicos y asociadas con la regulación epigenética, como la metilación, que puede silenciar la expresión génica.



**Figura 5.2:** Ejemplo de aplicación del juego del caos a una SG, dividiéndola por regiones. Se muestran en verde los puntos correspondientes a cada base de la secuencia que comienza con las bases CGTATTGC. El punto  $P_1$  se ubica a mitad de camino entre el punto inicial  $P_0 = (0.5, 0.5)$  y la esquina correspondiente a la base C  $= (0, 1)$ . El punto  $P_2$  se sitúa a mitad de camino entre el punto anterior  $P_1 = (0.25, 0.75)$  y la esquina correspondiente a la base G  $= (1, 1)$ , y así sucesivamente para las demás bases. Esta representación también crea una división en regiones determinadas por las últimas bases de la secuencia, de modo que cualquier secuencia que termine en C estará en el cuadrante correspondiente  $(0, 0.5) \times (0.5, 1)$ . Esta región puede subdividirse a su vez en cuatro regiones para secuencias que terminan en AC, CC, CG y CT, las cuales pueden dividirse nuevamente en las cuatro bases, aumentando la secuencia en un nucleótido.

Existen 64 combinaciones posibles de codones: 61 codifican aminoácidos específicos, como la metionina (ATG), que también actúa como codón de inicio para la traducción, y 3 son codones de parada (TAA, TAG, TGA). Estos últimos señalan el fin de la secuencia proteica, indicando al ribosoma que detenga la traducción. En la Figura 5.3 se muestra la ubicación de los puntos CGR por región, lo cual discutiremos en detalle más adelante. Esta distribución también permite localizar codones con sus aminoácidos correspondientes.

El papel de los codones es esencial, ya que establecen el marco de lectura de la secuencia genética y garantizan que la proteína se ensamble correctamente según la información genética almacenada en el ADN.

### 5.0.2. Representaciones Gráficas de Secuencias Genómicas

El primer método utilizado para obtener imágenes a partir de secuencias genómicas (SG) [6] asigna un punto dentro de un cuadrado a cada base de la SG, siguiendo el algoritmo del juego del caos, equivalente a un sistema de funciones iteradas (IFS) [2]. Este enfoque permite generar una imagen en el plano. Estos algoritmos dividen el cuadrado en zonas más pequeñas correspondientes a secuencias genómicas cortas específicas, facilitando el rastreo de la SG dentro de la imagen. La

|            |             |            |             |            |             |            |            |
|------------|-------------|------------|-------------|------------|-------------|------------|------------|
| CCC<br>pro | GCC<br>ala  | CGC<br>arg | GGC<br>gly  | CCG<br>pro | GCG<br>ala  | CGG<br>arg | GGG<br>gly |
| ACC<br>thr | TCC<br>ser  | AGC<br>ser | TGC<br>cys  | ACG<br>thr | TCG<br>ser  | AGG<br>arg | TGG<br>trp |
| CAC<br>his | GAC<br>asp  | CTC<br>leu | GTC<br>val  | CAG<br>glu | GAG<br>glu  | CTG<br>leu | GTG<br>Val |
| AAC<br>asn | TAC<br>try  | ATC<br>ile | TTC<br>phe  | AAG<br>lys | TAG<br>stop | ATG<br>Met | TTG<br>Leu |
| CCA<br>pro | GCA<br>ala  | CGA<br>arg | GGA<br>gly  | CCT<br>pro | GCT<br>ala  | CGT<br>arg | GGT<br>Gly |
| ACA<br>thr | TCA<br>ser  | AGA<br>arg | TGA<br>Stop | ACT<br>thr | TCT<br>ser  | AGT<br>ser | TGT<br>Cys |
| CAA<br>glu | GAA<br>glu  | CTA<br>leu | GTA<br>val  | CAT<br>his | GAT<br>Asp  | CTT<br>leu | GTT<br>val |
| AAA<br>Lys | TAA<br>Stop | ATA<br>ile | TTA<br>leu  | AAT<br>Asn | TAT<br>tyr  | ATT<br>ile | TTT<br>phe |

**Figura 5.3:** Zonas de correspondencia entre codones y aminoácidos (en símbolos) en el CGR, incluyendo codones de inicio y parada de genes. Los colores corresponden a las cuatro propiedades bioquímicas: no polar (amarillo), polar (verde), básico (púrpura) y ácido (rosa). Se observa que las regiones C y T son simétricas respecto a los aminoácidos, lo que significa que estas regiones corresponden exactamente a los mismos aminoácidos, mientras que las regiones A y G muestran ligera antisimetría debido a la presencia de codones de inicio y parada, manteniendo un equilibrio en las representaciones gráficas.

característica geométrica más importante es que la estructura resultante es autosimilar o fractal, lo cual surge del hecho de que la SG no sigue un patrón homogéneo ni en el orden ni en la cantidad de los diferentes tipos de bases.

## Representación del Juego del Caos

El CGR es un método iterativo que codifica secuencias biológicas en mapas fractales bidimensionales, ofreciendo un marco compacto e interpretable para la visualización genómica [23, 47–49]. En genómica, cada nucleótido (A, C, G, T) se asigna a una esquina de un cuadrado unitario, y los puntos se trazan marcando iterativamente el punto medio entre el punto anterior (inicialmente el centro del cuadrado) y el vértice correspondiente a la siguiente base (Figura 5.2). El patrón fractal resultante captura con alta fidelidad las frecuencias de k-méros y los sesgos de composición local [22, 50].

Dado que el CGR establece un mapeo biyectivo de k-méros a coordenadas fractales, preserva características biológicas esenciales y facilita comparaciones de secuencias libres de alineamiento, así como inferencia filogenética basada en métricas de distancia robustas [17, 51]. Su extensión basada en frecuencias (FCGR) convierte estos mapas en vectores de alta dimensión, que sirven como entradas efectivas para flujos de trabajo de aprendizaje automático, abarcando clustering, detección de anomalías y clasificación supervisada de elementos reguladores y genomas de patógenos.

Más allá de las representaciones estáticas, el CGR permite la exploración multiescala de paisajes

genómicos: ajustando la resolución de k-méros, los investigadores pueden “acercarse” para examinar motivos de secuencia específicos o “alejarse” para capturar características estructurales jerárquicas. Esfuerzos recientes han integrado el CGR en arquitecturas de aprendizaje profundo con mecanismos de atención, ejemplificado por el modelo CAPE, que integra la codificación de secuencias locales con conservación evolutiva para mejorar notablemente la predicción de la fuerza promotora [52].

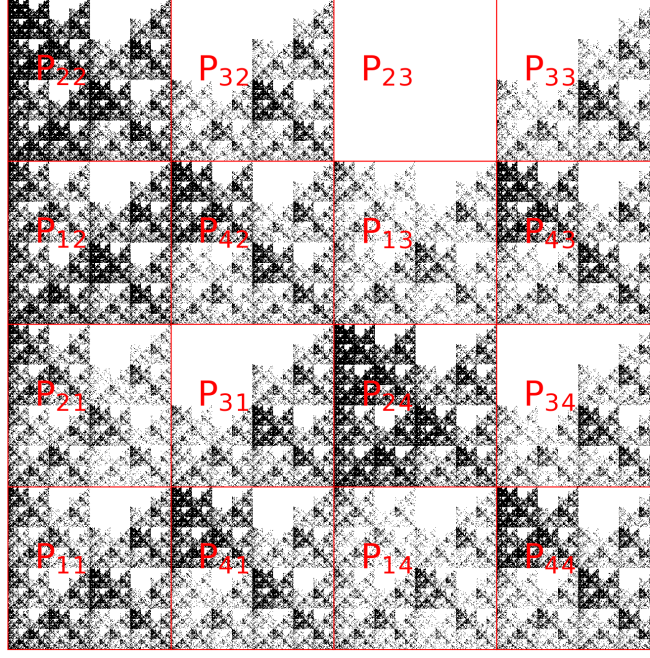
Estos avances subrayan la versatilidad del CGR: tiende un puente entre datos genómicos simbólicos y análisis numérico, mejora la interpretabilidad para expertos del dominio y escala eficientemente con conjuntos de datos de secuenciación a gran escala. Futuros trabajos podrían integrar el CGR con redes neuronales de grafos para modelar interacciones reguladoras de largo alcance (e.g., bucles promotor-potenciador) y extender los marcos basados en atención hacia la integración multi-ómica, ampliando así el impacto del CGR en genómica funcional y medicina de precisión.

El patrón obtenido de diferentes representaciones dependerá de la distribución de las bases en la SG. Independientemente de la longitud de la secuencia, es teóricamente posible rastrear en el CGR el origen y la secuencia que conduce a cada punto, ya que en todas las representaciones el espacio se divide en regiones que solo pueden ser accedidas mediante una secuencia dada (ver Figura 5.3), lo que significa que la SG puede representarse mediante un único número [16]. La limitación computacional que enfrentamos es que no podemos almacenar números infinitamente grandes o pequeños, por lo que es más conveniente seguir utilizando la SG original o el CGR.

La baja población o ausencia de una secuencia corta dentro de la SG resulta en regiones en la representación que no son visitadas [2]. Estas regiones se autorreplican a través de toda la CGR (ver Figura 5.4), permitiendo visualizar subsecuencias menos presentes, más presentes o ausentes en la SG.

Esta representación presenta desafíos al producir imágenes; dependiendo del número de puntos, las limitaciones en el tamaño del área de visualización y el tamaño de los puntos, la imagen tiende a saturarse para una longitud dada de la SG, dificultando distinguir el tipo de patrón obtenido (ver Figura 5.5). En algunos casos, es útil utilizar puntos opacos, pero esto no siempre funciona, ya que si la opacidad no se controla por regiones, la imagen se satura en áreas de alta concentración o desaparece en otras. Por esta razón, para esta y otras representaciones, dividimos el cuadrado en  $4^{12}$  celdas de igual tamaño y graficamos el número de puntos por celda (ver Figura 5.7).

Los CGR ocupan una cantidad significativa de memoria (aproximadamente 48 GB de almacenamiento para la secuencia genómica humana completa), lo que dificulta el almacenamiento y procesamiento en computadoras sin memoria suficiente. Por lo tanto, es ventajoso disponer de métodos de compresión eficientes que permitan analizar SGs muy grandes y que además aprovechen los patrones presentes en la SG, conduciendo a formas más simples de codificación.



**Figura 5.4:** Ejemplo de distribución de puntos generada usando MC. En la región  $P_{23}$ , la población es cero, lo cual se refleja en toda la representación. Este tipo de distribución ocurre cuando la probabilidad asociada con  $P_{23}$  en la matriz de Markov es cero.

Este procedimiento se expresa matemáticamente mediante la ecuación de iteración:

$$\begin{aligned} \vec{P}_n &= \frac{1}{2} \left( \vec{P}_{n-1} + \vec{V}_n \right), \quad \text{para } n > 0 \quad \text{con} \\ \vec{P}_0 &= \frac{1}{2} \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \quad \text{y} \\ \vec{V}_n &\in \left\{ \vec{A} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \vec{C} = \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \vec{G} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \vec{T} = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \right\}. \end{aligned} \quad (5.1)$$

Las coordenadas asignadas al  $n$ -ésimo nucleótido están dadas por:

$$\vec{P}_n = \frac{\vec{V}_n}{2} + \frac{\vec{V}_{n-1}}{2^2} + \cdots + \frac{\vec{V}_1}{2^{n-1}} + \frac{\vec{V}_0}{2^n} = \sum_{j=0}^{n-1} \frac{\vec{V}_j}{2^{n-j}}, \quad (5.2)$$

donde  $\vec{V}_0 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$ .

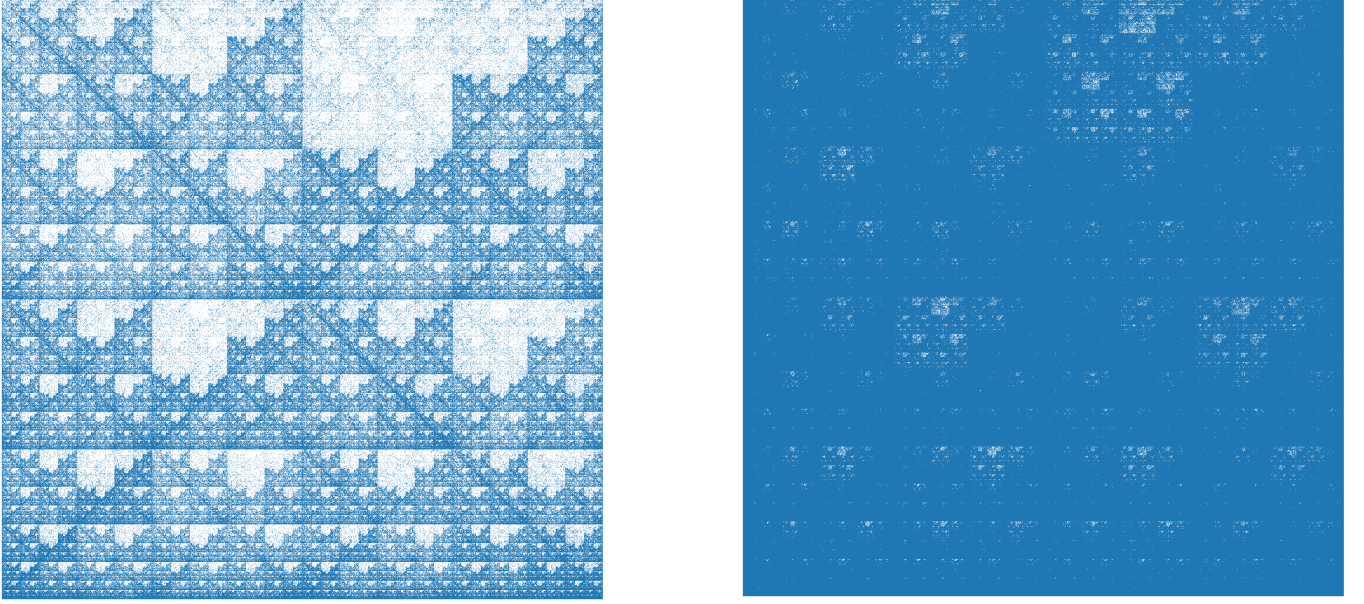
Una descripción matemática alternativa del CGR utiliza IFS, definido como el conjunto

$$\{w_A(\mathbf{x}), w_C(\mathbf{x}), w_G(\mathbf{x}), w_T(\mathbf{x})\}$$

con:

$$w_i \begin{pmatrix} x \\ y \end{pmatrix} = \frac{1}{2} \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{pmatrix} x \\ y \end{pmatrix} + \begin{pmatrix} V_{ix} \\ V_{iy} \end{pmatrix} = \frac{1}{2} \mathbb{I} \mathbf{x} + \vec{V}_i, \quad (5.3)$$





**Figura 5.5:** A la izquierda, imagen de un CGR con aproximadamente  $10^6$  puntos; a la derecha, con aproximadamente  $10^9$  puntos, donde el patrón ya no es distinguible.

para  $i \in \{A, C, G, T\}$ , donde

$$\vec{V}_A = \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \quad \vec{V}_C = \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \quad \vec{V}_G = \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \quad \vec{V}_T = \begin{pmatrix} 1 \\ 0 \end{pmatrix}. \quad (5.4)$$

Los puntos generados por la SG forman un caminante determinista (DW) con punto inicial  $\vec{P}_0$ :

$$\mathcal{O}(\vec{P}_0) = \left\{ \vec{P}_0, w_{V_1}(\vec{P}_0), w_{V_2} \circ w_{V_1}(\vec{P}_0), \dots, w_{V_N} \circ \dots \circ w_{V_1}(\vec{P}_0) \right\}. \quad (5.5)$$

El IFS asociado con el CGR determina el conjunto  $F$ , que es el soporte del fractal  $A$  generado por cualquier SG. Específicamente,  $F$  es el conjunto invariante bajo el operador de Hutchinson (HO):

$$\begin{aligned} W(F) &= w_A(F) \cup w_C(F) \cup w_G(F) \cup w_T(F) \\ &= \bigcup_i w_i(F) = F. \end{aligned}$$

En nuestro caso,  $F$  corresponde al cuadrado unitario abierto. Aplicando el HO  $K$  veces sobre  $Q$ , obtenemos una cobertura compuesta por  $4^K$  cuadrados de lado  $(1/2)^K$ . Las dimensiones de conteo de cajas y Hausdorff de  $F$  están dadas por:

$$D_B(F) = - \lim_{\varepsilon_n \rightarrow 0} \frac{\ln N_B(\varepsilon_n)}{\ln \varepsilon_n} = - \lim_{n \rightarrow \infty} \frac{\ln 4^n}{\ln 2^{-n}} = 2. \quad (5.6)$$

Por otro lado, la medida de Hausdorff  $H(F)$  y la dimensión  $D_H(F)$  se definen como:

$$H(F) = \lim_{\varepsilon_n \rightarrow 0} \sum_{n=1}^{N(\varepsilon_n)} \left( \frac{1}{2} \right)^{nD} = \lim_{n \rightarrow \infty} 4^n \left( \frac{1}{2^D} \right)^n = \lim_{n \rightarrow \infty} \left( \frac{4}{2^D} \right)^n;$$

$$H(F) = \begin{cases} \infty & \text{si } 4 > 2^D \Rightarrow D < 2, \\ 0 & \text{si } 4 < 2^D \Rightarrow D > 2. \end{cases} \quad \therefore \quad D_H(F) = 2;$$

Por tanto, la dimensión fractal de  $Q$  es 2. Sin embargo, la dimensión fractal de  $A$  se determinará utilizando el formalismo multifractal.

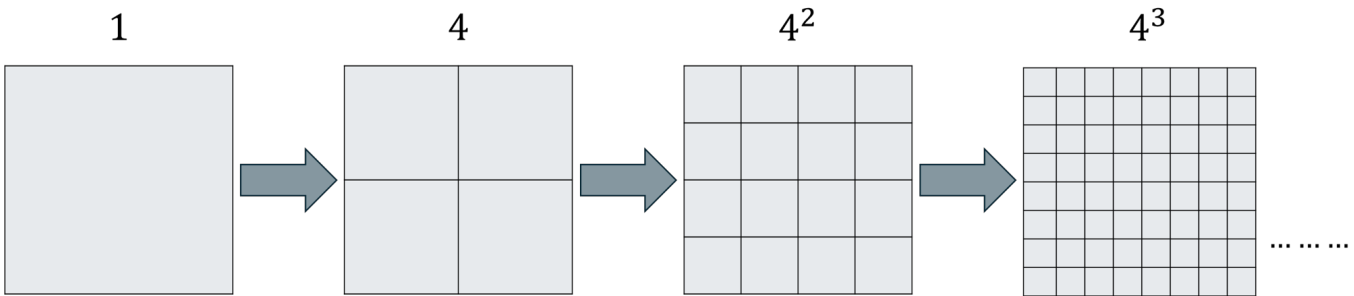
## Cadenas de Markov

En este método, construimos Cadenas de Markov (CM) asignando probabilidades de obtener cada una de las bases condicionadas por el estado anterior. Estos estados previos se definen como  $X$ , que son secuencias genómicas cortas de longitud fija  $n - 1$  para  $n = 1, 2, 3, \dots$ , y el estado final  $Y \in \{A, C, G, T\}$ .

Por ejemplo, para  $n = 2$ ,  $XY$  es un dinucleótido, es decir, una secuencia de dos bases donde  $X, Y \in \{A, C, G, T\}$ . Para  $n = 3$ ,  $X \in \{AA, AC, AG, AT, CA, CC, CG, \dots, TT\}$  corresponde a todos los posibles estados de dos bases, siempre con  $Y \in \{A, C, G, T\}$ , y de manera similar para otros valores de  $n$ . La matriz de probabilidades de la CM correspondiente dependerá de la longitud  $n$  y está dada por [4]:

$$P_{XY} = \frac{n_{XY}}{n_{XA} + n_{XC} + n_{XG} + n_{XT}}, \quad (5.7)$$

donde  $n_{XY}$  es el número de veces que  $XY$  aparece en la SG. Este proceso produce una matriz de probabilidades de tamaño  $4^{n-1} \times 4$ , generando una CM iterativamente al añadir una base en cada paso, dependiendo de las probabilidades correspondientes del estado anterior. La nueva SG resultante se utiliza para obtener un nuevo CGR. La Figura 5.4 muestra un ejemplo de un CGR generado usando una CM, donde la probabilidad  $P_{23} = 0$ .



**Figura 5.6:** División del cuadrado unitario utilizado en el método de conteo por cajas: Esta división se logra mediante la partición sucesiva del cuadrado en mitades verticales y horizontales. Cada paso produce un nuevo conjunto de cuadrados disjuntos de igual tamaño, cuya unión reconstruye el cuadrado original. Con cada iteración, el número total de cajas se cuadruplica.

Dado que estos son sistemas aleatorios originados a partir de la SG, la CM no permite una decodificación directa de vuelta a la SG, lo cual constituye su principal desventaja. Sin embargo, dependiendo de qué tan bien se ajuste a la distribución original, se puede realizar un análisis directo

sobre la matriz de probabilidades, que, dependiendo de  $n$ , puede tener significativamente menos elementos que la SG. En este trabajo, consideraremos solo cuatro valores diferentes de  $n$  (2, 3, 6, 12), por lo que el número máximo de elementos en las matrices de probabilidades es  $4^{12} = 16777216$ .

Otro aspecto a considerar en la CM es su longitud, asegurando que sea lo suficientemente grande para generar una distribución cercana a la original; en otras palabras, qué tan rápido alcanzamos el atractor, asumiendo que la distribución de bases en la SG se sitúa sobre un soporte fractal.

### Representación Genómica Binaria

Inspirados por las ideas propuestas por [16], proponemos un nuevo método para comprimir y representar SGs, que consiste en asignar vectores binarios bidimensionales a cada una de las bases:

$$A = \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \quad C = \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \quad G = \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \quad T = \begin{pmatrix} 1 \\ 0 \end{pmatrix}. \quad (5.8)$$

De esta manera, cada punto en el plano se representa en forma binaria, y los puntos en la Representación Genómica Binaria (BGR) se definen como:

$$P_N = 0.V_N V_{N-1} V_{N-2} \dots V_2 V_1 V_0, \quad (5.9)$$

donde  $V_j \in \{A, C, G, T\}$ . Este punto incluye el punto anterior en la secuencia, ya que:

$$P_{N-1} = 0.V_{N-1} V_{N-2} \dots V_2 V_1 V_0, \quad (5.10)$$

Por lo tanto, los puntos en la representación se obtienen desplazando el punto binario hacia la derecha. Así, la BGR nos permite almacenar la SG como dos secuencias de ceros y unos. Es evidente que siempre es posible revertir a la SG original.

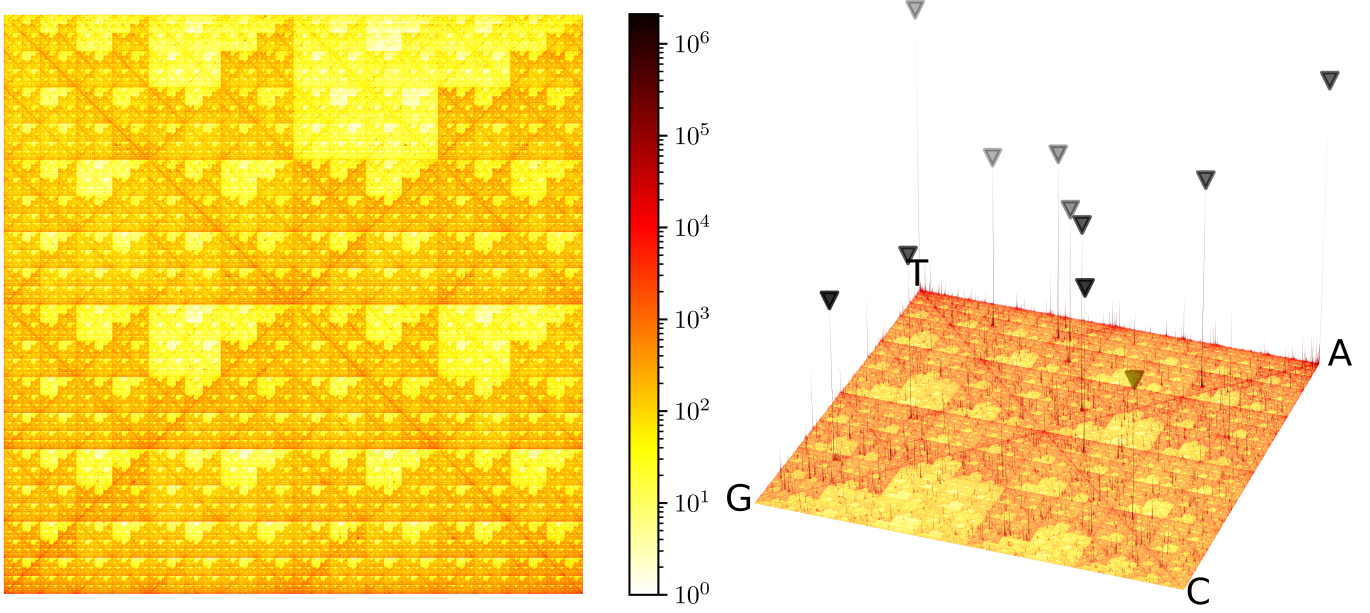
La forma en que se codifica la representación binaria es mediante la segmentación de la SG; cada uno de los segmentos de longitud  $M$  representa un punto en el plano, por lo que la SG generalmente se representa por  $n + 1 = \left\lfloor \frac{\text{número de bases en la SG}}{M} \right\rfloor + 1$  puntos.

La versatilidad de la BGR permite asociar una SG con  $N$  puntos dentro o fuera del cuadrado unitario  $Q$  (ver Fig. 5.2). Esto es posible porque la información contenida en la Ecuación 5.2 es equivalente al vector entero:

$$\vec{I}_N = 2^0 \vec{V}_0 + 2^1 \vec{V}_1 + \dots + 2^N \vec{V}_N = \sum_{i=0}^N 2^i \vec{V}_i = \vec{V}_0 \vec{V}_1 \dots \vec{V}_N,$$

donde la notación  $\vec{V}_0 \vec{V}_1 \dots \vec{V}_N$  representa la concatenación de vectores.

Al caracterizar la SG usando  $\vec{I}_N$ , empleamos la forma binaria de la Representación del Juego del Caos Entera (iCGR) de Yin. Toda la SG se define mediante tres enteros:  $(N, I_{Nx}, I_{Ny})$ . La iCGR codifica la SG sin pérdida de información usando  $p+1$  puntos en lugar de los  $N$  requeridos por el CGR



**Figura 5.7:** Gráfica en escala logarítmica del número de puntos por celda en el CGR, construida utilizando la metodología descrita en la Sección 5.0.2 para la SG CHM13v2.0, en representaciones 2D y 3D. La representación bidimensional y tridimensional consta de un total de  $4^{12}$  celdas. Los triángulos negros en la figura 3D de la derecha marcan algunos de los valores máximos en la distribución. Estas imágenes muestran la estructura fractal de la distribución en el CGR.

estándar. Esta codificación permite la representación gráfica de la SG completa. Específicamente,  $\vec{I}_N$  puede descomponerse como:

$$\vec{I} = \sum_{j=0}^N \vec{V}_j 2^j = \sum_{\alpha=0}^{p-1} 2^{\alpha M} \vec{N}_{\alpha} + 2^{pM} \vec{N}_p,$$

$$\text{con } p = \left\lfloor \frac{N+1}{M} \right\rfloor, \quad q = (N+1) \bmod M, \quad (5.11)$$

donde:

$$\vec{N}_{\alpha} = \sum_{k=0}^{M-1} \vec{V}_{\alpha M+k} 2^k; \quad \vec{N}_p = \sum_{k=0}^q \vec{V}_{pM+k} 2^k.$$

La SG se caracteriza por  $p+1$  vectores enteros:

$$\begin{pmatrix} N_{0X} \\ N_{0Y} \end{pmatrix}, \begin{pmatrix} N_{1X} \\ N_{1Y} \end{pmatrix}, \dots, \begin{pmatrix} N_{(p-1)X} \\ N_{(p-1)Y} \end{pmatrix}, \begin{pmatrix} N_{pX} \\ N_{pY} \end{pmatrix}.$$

Cada vector corresponde a  $M$  nucleótidos, excepto el último, que corresponde a  $q$ . Al expresar estos vectores en binario y concatenarlos, reconstruimos el vector entero  $\vec{I}_N$  asociado con la SG completa.

### 5.0.3. Formalismo Multifractal

En esta sección, discutimos el método utilizado para obtener la caracterización estadística de la SG en términos del formalismo multifractal [11, 24]. Comenzamos dividiendo  $Q$  en  $4^k$  celdas de lado  $\left(\frac{1}{2}\right)^k$  y determinando con qué frecuencia el caminante determinista (DW) generado por la SG visita

cada celda. Esto produce el vector de probabilidad:

$$\vec{p} = (p_1, p_2, \dots, p_R), \quad (5.12)$$

donde  $R = 4^k$ . La probabilidad de escolta de orden  $q$  se define como [25]:

$$\pi_i(q) = \frac{p_i^q}{Z(q)}, \quad \text{donde } Z(q) = \sum_{i=1}^R p_i^q. \quad (5.13)$$

La información de Rényi de orden  $q$  está dada por [53]:

$$I_q(\vec{p}) = \frac{1}{q-1} \ln Z(q). \quad (5.14)$$

La función de partición  $Z(q)$  satisface la ley de potencias:

$$Z(q) \approx \Delta^{\tau(q)}, \quad (5.15)$$

donde  $\Delta$  denota el tamaño de la celda, y  $\tau(q)$ , llamado exponente de masa, se define como:

$$\tau(q) = \frac{\ln Z(q)}{\ln \Delta}. \quad (5.16)$$

Usando (5.14) y (5.16), se deriva la dimensión generalizada de Rényi de orden  $q$ :

$$D_q = \frac{I_q(\vec{p})}{\ln \Delta} = \frac{1}{q-1} \frac{\ln Z(q)}{\ln \Delta} = \frac{\tau(q)}{q-1}. \quad (5.17)$$

La intensidad estadística de cada celda se cuantifica mediante el exponente de Hölder  $\alpha_i$  [9, 25]:

$$\alpha_i = \frac{\ln \mu_i(\Delta)}{\ln \Delta} = \frac{\ln p_i}{\ln \Delta}, \quad \text{para } i = 1, 2, \dots, R, \quad (5.18)$$

donde  $\mu_i(\Delta)$  denota la medida estadística de la  $i$ -ésima celda. La función de partición puede expresarse en términos de los exponentes de Hölder como:

$$Z(q, \Delta) = \sum_{i=1}^R \Delta^{\alpha_i}. \quad (5.19)$$

Para  $\Delta$  pequeño, el espectro multifractal  $f(\alpha)$  se aproxima por:

$$f(\alpha) \approx -\frac{\ln N(\alpha)}{\ln \Delta}, \quad (5.20)$$

donde  $N(\alpha)$  es el número de celdas con  $\alpha < \alpha_i < \alpha + d\alpha$ . La función de partición se convierte entonces en:

$$Z(q, \Delta) = \sum_{\alpha} \Delta^{q\alpha - f(\alpha)} d\alpha. \quad (5.21)$$

Cuando  $\Delta \rightarrow 0$ , la contribución dominante proviene del valor  $\alpha^*$  que minimiza la función:

$$g(\alpha) = q\alpha - f(\alpha),$$

determinado por las condiciones:

$$\left. \frac{df(\alpha)}{d\alpha} \right|_{\alpha^*} = q; \quad \left. \frac{d^2 f(\alpha)}{d\alpha^2} \right|_{\alpha^*} < 0. \quad (5.22)$$

Así, para  $\Delta$  pequeño, obtenemos:

$$Z(q, \Delta) \sim \Delta^{g(\alpha^*)} = \Delta^{q\alpha^* - f(\alpha^*)}. \quad (5.23)$$

Combinando este resultado con (5.16) se establece una conexión entre el exponente de masa y la dimensión fractal del conjunto donde la medida estadística se concentra:

$$\tau(q) = q\alpha^* - f(\alpha^*). \quad (5.24)$$

Tomando la diferencial de (5.18) y usando (5.21), derivamos:

$$d\tau = q d\alpha^* + \alpha^* dq - df(\alpha^*) = \alpha^* dq, \quad (5.25)$$

demostrando que  $\tau(q)$  y  $f(\alpha^*)$  son transformadas de Legendre una de la otra. Ambas funciones codifican información equivalente.

Para  $\Delta$  pequeño, se cumplen las siguientes aproximaciones:

$$\alpha^*(q) = \frac{d\tau}{dq} \approx \frac{d}{dq} \left( \frac{\ln Z(q)}{\ln \Delta} \right) = \sum_{i=1}^R \pi_i(q) \frac{\ln p_i}{\ln \Delta}; \quad (5.26)$$

$$f(q) = q\alpha^*(q) - \tau(q) \approx \frac{1}{\ln \Delta} \sum_{i=1}^R \pi_i(q) \ln \pi_i(q). \quad (5.27)$$

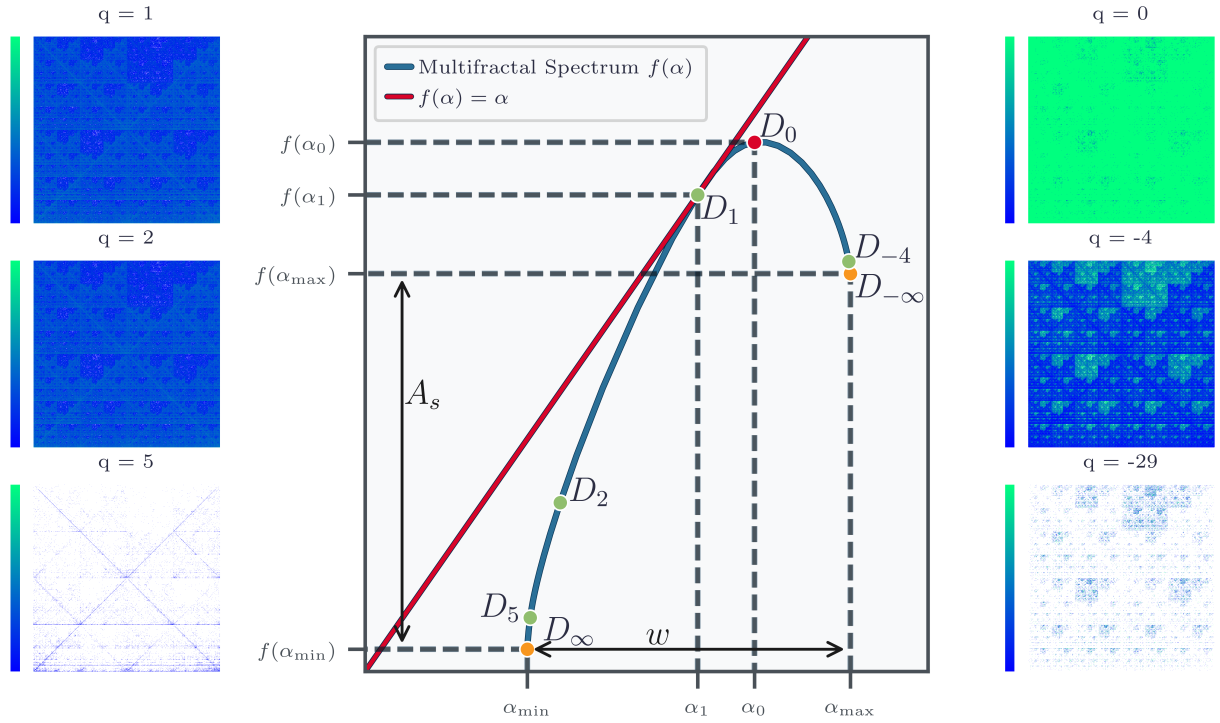
El espectro multifractal se obtiene eliminando el parámetro  $q$  de (5.20) y (5.21) [11].

Enfatizamos que  $\alpha(q)$  representa el exponente de Hölder promedio ponderado por la distribución de escolta, mientras que  $f(q)$  corresponde al exponente de escalamiento de la entropía de Shannon de esta distribución. Esto demuestra que la información multifractal está completamente encapsulada en la distribución de escolta.

#### 5.0.4. Análisis Multifractal

Las diferentes representaciones de la SG sirven como herramientas fundamentales para la codificación y decodificación. Estas exhiben características cualitativas—como autosimilitud y distribución sobre soportes geométricos—que confieren una naturaleza multifractal a la SG [24]. Aunque no existe una definición única de fractalidad, generalmente se acepta que a estos conjuntos se les puede asignar una dimensión no entera, que puede medirse mediante varios métodos [2]. En este estudio, empleamos el método de conteo de cajas dentro del marco del CGR para realizar un análisis multifractal. Este enfoque es particularmente conveniente porque el CGR restringe el análisis a un área cuadrada





**Figura 5.8:** Ejemplo del espectro multifractal  $f(\alpha)$  y las ubicaciones correspondientes de las dimensiones generalizadas para valores de  $q$  (-29, -4, 0, 1, 2, 5), junto con sus distribuciones asociadas. Cada imagen a los lados ilustra el cambio en la distribución producido al variar  $q$ —con valores positivos correspondientes a altas frecuencias y valores negativos que reflejan mejor las bajas frecuencias. La figura demuestra además la transición entre regímenes de probabilidad extrema en el CGR y destaca componentes clave del espectro multifractal, incluyendo la dimensión fractal  $D_0$  (que corresponde al máximo de cualquier espectro), el ancho  $w$  (que cuantifica la heterogeneidad multifractal del objeto) y la asimetría  $A_s$  (generalmente asociada con la ocurrencia de eventos raros, determinada por la extensión de las regiones que cubren).

definida. El método de construcción de la SG basado en CGR asigna una región cuadrada a cada secuencia posible según su longitud. Por ejemplo, las secuencias de 12 nucleótidos se distribuyen entre  $4^{12}$  regiones, lo que corresponde exactamente al número de secuencias distintas que pueden generarse a partir de cuatro nucleótidos sobre una longitud de 12. Cada vez que se lee una secuencia (de acuerdo con la longitud especificada), el CGR asigna un punto a la región correspondiente. La acumulación de estos puntos en diferentes regiones—normalizada por el número total de puntos (es decir, la longitud total de la secuencia)—produce una densidad de probabilidad, como se muestra en las Figuras 5.7 y 5.8.

La dimensión fractal estándar ( $D_0$ ), también conocida como dimensión de Hausdorff-Besicovitch, es una medida topológica que cuantifica solo la ocupación espacial de un objeto fractal, sin considerar las heterogeneidades en la distribución de frecuencias. Este enfoque asume implícitamente una homogeneidad idealizada, donde las regiones ocupadas exhiben densidad uniforme y los vacíos se distribuyen aleatoriamente. Formalmente,  $D_0$  se deriva mediante el método de conteo de cajas, que determina el número mínimo de cajas  $N(r)$  de tamaño  $r$  requeridas para cubrir el soporte fractal  $F$ :

$$D_0 = \lim_{r \rightarrow 0} \frac{\log N(r)}{\log(1/r)}. \quad (5.28)$$

---

En el plano euclidiano  $\mathbb{R}^2$ ,  $D_0$  está acotada entre 0 y 2:

- $D_0 \rightarrow 0$ : Corresponde a soportes discretos (puntos aislados, conjuntos numerables) o estructuras completamente desconectadas.
- $D_0 = 2$ : Describe conjuntos densos en  $\mathbb{R}^2$  con autosimilitud exacta, como la curva de Peano.

Si bien es efectiva para clasificar fractales clásicos, esta métrica no logra caracterizar sistemas complejos con distribuciones multifractales, donde coexisten regiones de alta y baja densidad. La generalización a dimensiones multifractales ( $D_q$ ) mediante el formalismo de Rényi aborda esta limitación al incorporar momentos estadísticos de orden  $q$  [3].

Las relaciones utilizadas en este análisis para obtener el espectro se basan en dos exponentes asociados con la frecuencia y el soporte geométrico:  $q$  y  $\tau$ , respectivamente. Estos exponentes se relacionan mediante la siguiente ecuación:

$$\tau(q) = \lim_{r \rightarrow 0} \frac{\log \sum_{l=1}^N p_l^q}{\log r}, \quad (5.29)$$

donde  $N$  es el número de cajas, y  $p_l = \frac{n_l}{N_{\text{total}}}$  representa la probabilidad de ocupación normalizada (o masa) en la  $l$ -ésima caja. Aquí,  $N_{\text{total}}$  denota el número total de puntos en la secuencia y  $n_l$  el número de puntos en esa caja.

El exponente  $q$  está relacionado con los momentos estadísticos, y las variaciones en su valor conducen a cambios en la distribución y los pesos estadísticos dentro del CGR. La Figura 5.8 muestra un ejemplo de cómo la distribución del CGR varía para diferentes valores de  $q$ . Estas representaciones también se conocen como conjuntos de condensación. Para valores de  $q$  positivos grandes, dominan las regiones de alta probabilidad; estas regiones están asociadas con características fuertes o dominantes. En el caso de la SG, esto corresponde a secuencias cortas dentro de una SG de longitud  $n$  que exhiben las frecuencias más altas en el CGR—es decir, las secuencias más recurrentes en la SG. Estas regiones de alta medida tienden a ser finas u ocupar un área pequeña, apareciendo como líneas (como se observa en secuencias compuestas por los nucleótidos AT o GT) o como grupos de puntos dispersos generalmente asociados con regiones repetitivas en la SG, con contribuciones menores de otras áreas.

Para valores de  $q$  negativos, se acentúan las regiones de baja probabilidad. En la SG humana, el enfoque se desplaza hacia eventos raros o extremos—es decir, secuencias cortas con frecuencias bajas, como aquellas que terminan en el dinucleótido CG o lo contienen. Dado que este dinucleótido rara vez se repite, genera los espacios vacíos en el CGR, que se replican de manera autosimilar a lo largo de toda la representación.

Es bien conocido que los momentos estadísticos están estrechamente ligados a la forma de una distribución. Por ejemplo, en una distribución normal (gaussiana), el segundo momento es el primer momento no nulo, responsable de su forma característica de campana. Sin embargo, en el contexto



multifractal, confiar en un solo momento es insuficiente; se debe considerar un rango de valores de  $q$  para capturar con precisión la forma verdadera de la distribución.

El exponente  $\tau(q)$ , también conocido como exponente de masa, describe cómo cambia la distribución con la escala (ver Ecuación 5.29). En otras palabras, a medida que varía la escala, se hace evidente una concentración de puntos en el CGR. Para una distribución homogénea—como una secuencia que repite consistentemente el mismo nucleótido, indicando una estructura repetida continuamente—se observaría una variación lineal de  $\tau(q)$  con respecto a  $q$ , caracterizando un conjunto monofractal. Por el contrario, las estructuras multifractales exhiben un comportamiento no lineal, indicando que la secuencia cambia de manera discontinua y heterogénea, combinando diferentes tipos de patrones. Esto, a su vez, refleja el grado de multifractalidad de la secuencia: cuanto mayor es la variación en  $\tau(q)$ , más compleja es su estructura multifractal.

Otra cantidad útil para el análisis es la dimensión de Rényi o dimensión generalizada [3]:

$$D_q = \frac{\tau(q)}{q-1}, \quad (5.30)$$

que contiene la misma información que  $\tau(q)$  pero asigna una dimensión fractal a cada valor de  $q$ .

También es importante notar que, dependiendo del objeto de análisis, cada momento puede tener interpretaciones diferentes; en consecuencia, la dimensión fractal está inherentemente vinculada al momento específico bajo consideración. Por ejemplo,  $q = 0$  está asociado únicamente con la distribución espacial de puntos, independientemente de su concentración. En secuencias genómicas, este valor simplemente indica qué patrones pueden ocurrir y cuáles no, estableciendo así el soporte fractal para cualquier distribución. La dimensión correspondiente a  $q = 1$ , denotada como  $D_1$ , se conoce como dimensión de información y representa la dimensión fractal de la distribución de probabilidad. Esta dimensión a menudo se interpreta en términos de la incertidumbre o predictibilidad de un sistema; para una secuencia dada, refleja la probabilidad de que una secuencia sea seguida por una secuencia subsiguiente específica. Dado que las secuencias se leen nucleótido por nucleótido, esta interpretación esencialmente se reduce a cambios en el nucleótido final, como se describe en los modelos de cadenas de Markov.

Además, las distribuciones fractales tienden hacia un límite o atractor; es decir, a medida que se modifica el rango de escalas o se consideran valores de  $r$  más pequeños, la distribución converge a una forma fija—representando el estado de equilibrio del sistema. En consecuencia, la medida asociada aumenta hasta cierto punto, caracterizando la tasa de crecimiento de la entropía de Shannon [3, 24]. Otra dimensión comúnmente discutida es la dimensión de correlación  $D_2$ , que cuantifica cómo se agrupan diferentes puntos o subconjuntos, reflejando la variación conjunta entre los componentes fractales. Así, en los CGRs obtenidos a partir de SGs, las regiones de acumulación se localizan predominantemente a lo largo de los bordes—donde ocurren secuencias compuestas únicamente de A's y T's o exclusivamente de T's y G's—y también a lo largo de las diagonales, correspondientes

a secuencias compuestas de A's y G's o T's y C's. Esto se debe a que las distancias probabilísticas en estas regiones son generalmente más cortas, alineándose con las que normalmente se encuentran dentro de la distribución de probabilidad promedio. Las dimensiones generalizadas satisfacen la relación de orden  $D_{q'} \geq D_q$  para  $q' < q$ . Además, en sistemas multifractales, estas dimensiones exhiben un comportamiento asintótico cuando  $q$  se aproxima a  $\pm\infty$ , mientras que en conjuntos homogéneos, todas las dimensiones generalizadas son iguales.

Finalmente, el espectro multifractal se obtiene mediante la transformada de Legendre:

$$f(\alpha) = \inf_q (q\alpha - \tau(q)), \quad (5.31)$$

donde se cumplen las siguientes relaciones:

$$\alpha = \frac{d\tau}{dq}, \quad q = \frac{df}{d\alpha}. \quad (5.32)$$

El valor  $\alpha$  se conoce como exponente de Hölder (o de singularidad) y está directamente vinculado a la medida del sistema. En el CGR, por ejemplo,  $\alpha$  cuantifica la concentración de puntos en una escala dada, que, en el contexto de la SG, corresponde a la frecuencia de secuencias de nucleótidos de longitud  $n$ . Al disminuir la escala, se considera un número mayor de nucleótidos; sin embargo, dado que la SG es finita, la escala no puede reducirse indefinidamente—una limitación inherente a todos los escenarios del mundo real y prácticas numéricas. En otras palabras, a cada secuencia se le puede asignar una medida que depende de la escala. Además,  $\alpha$  representa la tasa de cambio del exponente  $\tau(q)$  (ver Ecuación 5.32), que caracteriza el comportamiento de los sistemas fractales. En un sistema completamente homogéneo,  $\tau(q)$  aumenta a una tasa constante y  $\alpha$  permanece constante, lo que significa que los fractales simples exhiben un único valor de  $\alpha$ . Por el contrario, los sistemas multifractales muestran un rango de valores de  $\alpha$ ; la extensión de este rango indica el grado de multifractalidad o complejidad estructural del sistema. Así, el espectro multifractal  $f(\alpha)$  revela no solo si un objeto es multifractal, sino también en qué medida exhibe multifractalidad, permitiendo así una caracterización más precisa de imágenes o comportamientos fractales.

En la Figura 5.8, se proporciona un ejemplo que ilustra la forma típica y las características clave del espectro multifractal  $f(\alpha)$  para SGs. Este esquema destaca los valores clave del espectro y sus dimensiones fractales correspondientes. La dimensión convencional  $D_0$ , que representa la dimensión de Hausdorff del soporte fractal, marca el valor máximo del espectro. A cada lado, los valores  $q > 0$  (izquierda) y  $q < 0$  (derecha) diferencian entre regiones de alta y baja probabilidad, respectivamente, dentro de la SG.

En el ejemplo mostrado, se observa una mayor variabilidad en las regiones de alta probabilidad (asociadas con  $q > 0$ ), lo que sugiere que las regiones de baja probabilidad (por ejemplo, secuencias raras o únicas) exhiben una mayor estabilidad en el CGR. Estas últimas se desvían menos del soporte fractal estándar ( $D_0$ ), mientras que las regiones de alta probabilidad (por ejemplo, dominios

conservados) muestran fluctuaciones significativas. La diferencia entre estos comportamientos puede cuantificarse utilizando la asimetría  $A_s$ , una métrica que refleja la propensión de la secuencia a contener eventos extremos o raros.

## 5.1. Metodología

Este estudio utiliza exclusivamente datos de la SG CHM13 y la información proporcionada en [8]. La metodología se centra principalmente en generar las representaciones gráficas descritas en la Sección 5.0.2, que destacan la estructura multifractal de la SG tanto para el genoma completo como para su partición en los 24 cromosomas y el ADN mitocondrial.

El análisis facilita la identificación de distribuciones de puntos—u ocupación—dentro del CGR, subdividido en celdas. Este procedimiento, conocido como conteo por cajas [2], implica particionar recursivamente el cuadrado que contiene el CGR en cuadrados más pequeños y no superpuestos (ver Figura 5.6). A cada cuadrado se le asigna entonces un valor que representa el número de puntos del CGR que contiene. En este estudio, se aplicaron subdivisiones en  $4^{12}$  y  $4^{10}$  cajas para evaluar las distribuciones de puntos en los CGRs de la SG.

Estas distribuciones se emplearon posteriormente en el análisis multifractal para comparar e identificar diferencias y similitudes cuantitativas dentro de la SG. Para este propósito, se calcularon los valores  $p_i$  para cada caja, y se derivó el exponente de masa  $\tau$  usando la Ecuación 5.29 para valores de  $q$  en el rango de  $-30$  a  $30$ . Adicionalmente, se determinaron las dimensiones fractales y el espectro multifractal usando las Ecuaciones (5.30) y (5.31).

También generamos distribuciones unidimensionales de CGR, como se muestra en la Figura 5.10, que son equivalentes a las representaciones originales. Este enfoque nos permite consolidar los resultados obtenidos individualmente para el ensamblaje completo, cromosomas individuales y ADN mitocondrial, facilitando así la identificación aproximada de diferencias en las distribuciones y las regiones donde ocurren.

En la Sección 5.0.1, describimos la metodología utilizada para derivar una visualización unidimensional alternativa del CGR, en la cual cada base se representa como parte de una serie temporal (ver Figura 5.15). Para este propósito, cada punto en el CGR se transforma de modo que se distribuya a lo largo de una sola dimensión mientras también indica su posición en la SG. El único objetivo de esta transformación es demostrar que el CGR también puede usarse para identificar cambios estructurales que se manifiestan como bandas a lo largo de toda la SG y en todos los cromosomas.

Aplicamos los métodos de CM y BGR a la SG humana completa para obtener la matriz de probabilidad correspondiente y el iCGR, respectivamente (ver Sección 5.0.2). Usando la matriz de probabilidad, generamos una secuencia aleatoria de bases de la misma longitud que la SG, aplicamos el método CGR a esta secuencia y obtuvimos su distribución de puntos correspondiente. La BGR

genera directamente una distribución de puntos similar a la obtenida con el CGR, revelando así la distribución asociada. Estos resultados se analizaron multifractalmente de la misma manera que para el ensamblaje y los cromosomas, permitiendo una comparación posterior con los resultados obtenidos del CGR tradicional.

Finalmente, utilizamos los resultados del CGR para demostrar una aplicación del método para identificar la arquitectura genómica en los cromosomas 9 e Y. Con este fin, identificamos los valores máximos en la distribución para cada uno de estos cromosomas, correspondientes a secuencias específicas basadas en las reglas de partición por cajas del CGR (figuras 5.2 y 5.3). Una vez logrado esto, observamos qué regiones de la secuencia corresponden a—y están típicamente asociadas con—la estructura, características y funciones biológicas de la SG.

## 5.2. Resultados y Análisis

Se obtuvieron imágenes del CGR para la SG del ensamblaje T2T-CHM13v2.0, así como para cromosomas individuales y ADN mitocondrial (ver Figuras 5.7 y 5.9). Para mantener la estructura fractal y considerar el predominio de ciertas regiones, se empleó una escala logarítmica. Aunque el ADN mitocondrial no forma parte del ensamblaje y no es el foco principal de este estudio, lo incluimos para ilustrar que su estructura multifractal es claramente diferente, exhibiendo una mayor presencia de la base C, lo que hace que su soporte se asemeje a un triángulo de Sierpinski [2].

Dado que la longitud de la secuencia de ADN mitocondrial es significativamente menor que la de los cromosomas, es apropiado considerar un análisis con menos cubrimientos. Basándonos en el comportamiento de la distribución respecto al número de cajas y las limitaciones computacionales, encontramos que a partir de  $4^{10}$  se obtiene una buena aproximación al atractor de la distribución.

Respecto a la distribución de bases dentro de las secuencias, observamos un número ligeramente mayor de bases A y T, aunque la diferencia no es desproporcionada en ninguno de los cromosomas. En términos de dinucleótidos, el CGR muestra claramente que la región correspondiente a CG tiene la menor cantidad de puntos, y esta ausencia se refleja en todo el CGR, particularmente dentro de sí misma. Esta característica sugiere una probabilidad significativamente reducida de que la base C sea seguida inmediatamente por la base G. Esta comprensión de la secuenciación se visualizará más fácilmente a través de las CM.

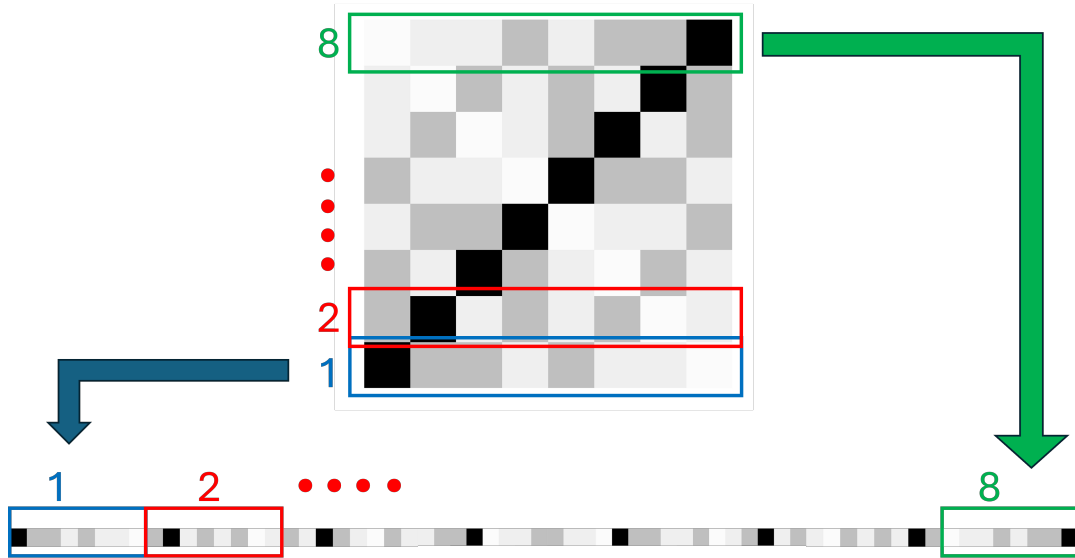
La escasez de dinucleótidos CG en la representación ya ha sido reportada en otros vertebrados [21], junto con la existencia de islas CG o GC, sugiriendo que esta representación puede servir como una firma genómica para varias especies. Adicionalmente, es notable que los picos significativos en la distribución, compuestos principalmente por bases A y T en la mayoría de los cromosomas, generalmente corresponden a secuencias repetitivas largas dentro del genoma, indicando una clara separación entre regiones codificantes y no codificantes.



**Figura 5.9:** Gráficas del número de puntos por celda en el CGR para cada cromosoma y el ADN mitocondrial. En cada caso, las frecuencias corresponden a los conteos dentro de las  $4^{10}$  celdas. Se aplicó una escala logarítmica a todas las visualizaciones.

En la Figura 5.11, generamos un arreglo unidimensional para los CGRs tanto del ensamblaje completo como de los cromosomas individuales (siguiendo la metodología descrita en el diagrama presentado en la Figura 5.10). Calculamos la densidad  $\rho$  (el número de puntos en cada celda relativo a la longitud de la respectiva secuencia genómica) y la densidad máxima  $\rho_{\text{máx}}$  (el número de puntos en cada celda comparado con la celda con el mayor conteo de puntos en el CGR correspondiente). Este proceso de normalización ajusta todas las representaciones a máximos unitarios. Esta visualización también demuestra claramente que el CGR total es esencialmente una superposición de los CGRs cromosómicos individuales.

En la Figura 5.12, se muestran las gráficas del exponente  $\tau$ , la dimensión generalizada y el espectro multifractal. Aunque todos contienen casi la misma información, decidimos presentarlos juntos con la



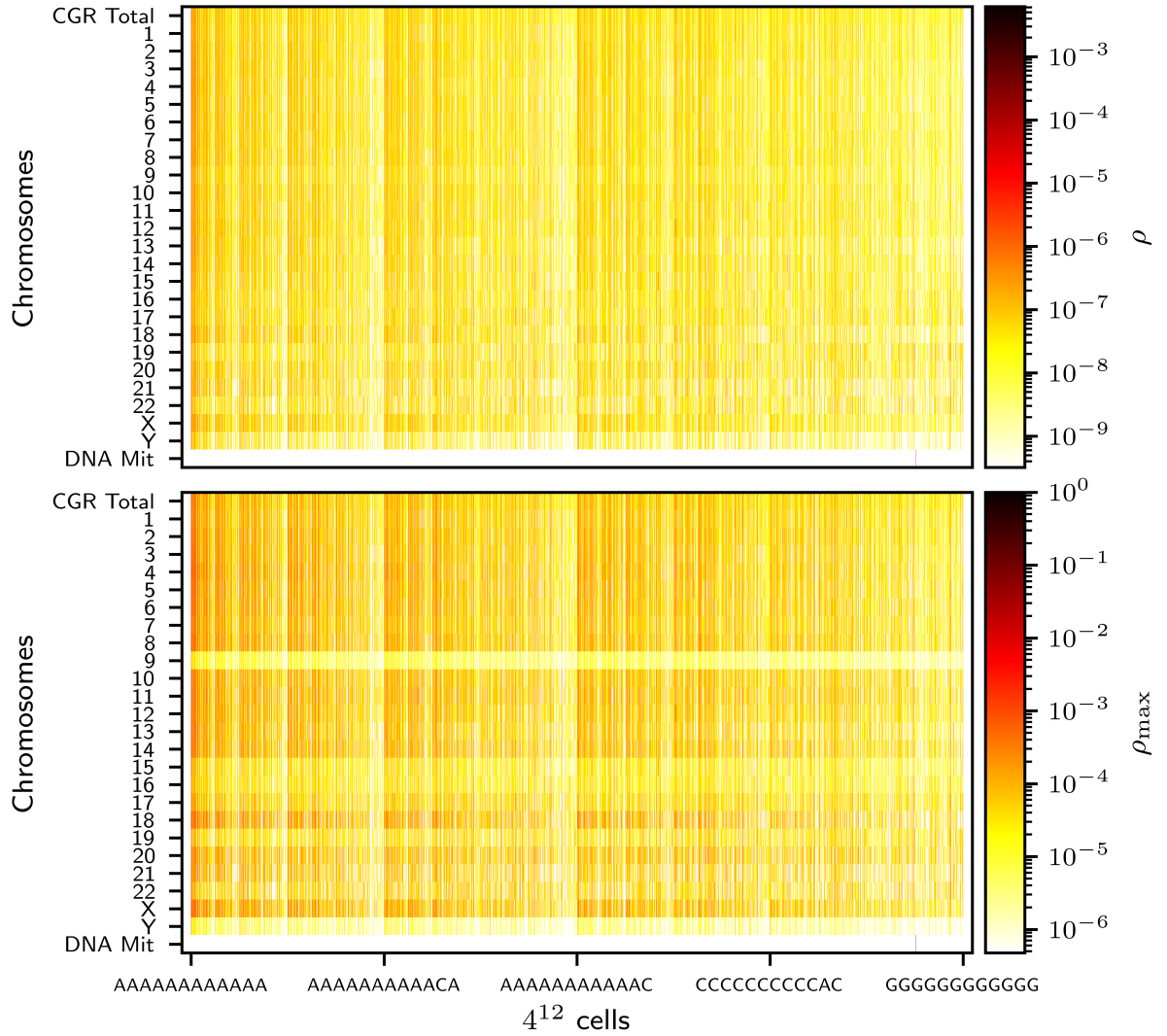
**Figura 5.10:** Ejemplo de conversión de una representación bidimensional a una unidimensional. En este caso, muestra la distribución de un CGR dividido en  $4^3$  cubrimientos cuadrados, donde la intensidad del color en cada cuadrante refleja la medida correspondiente en esa región. Para la conversión, tomamos cada fila, de abajo hacia arriba, y las alineamos a lo largo de una línea. Este procedimiento se aplica rutinariamente a CGRs obtenidos a partir de la SG, que se dividen en  $4^{12}$  cubrimientos y  $4^6$  filas.

intención de lograr una mayor distinción entre los diferentes resultados. Sin embargo, nos centraremos en el espectro multifractal, para el cual proporcionamos un resumen de cada espectro.

Claramente, las medidas se ven significativamente afectadas porque la longitud de la SG correspondiente al ensamblaje es mucho mayor que las longitudes de los cromosomas. Para corregir aproximadamente esta disparidad, consideramos un número menor de cubrimientos para las SGs cromosómicas. La Figura 5.13 muestra los resultados, donde se puede observar un mejor ajuste en el área alrededor de  $\alpha = 2$ , que es el área de mayor interés, aunque con pérdidas en los extremos.

De los resultados, podemos destacar que aunque el comportamiento general de los cromosomas es similar—centrado alrededor de la dimensión fractal estándar  $D_0$  (como se observa más claramente en la Figura 5.13)—el análisis detallado presentado en la Figura 5.12 revela que el espectro multifractal de cada cromosoma es distinto, dispersándose más en sus extremos. En particular, los cromosomas 9 e Y exhiben un claro sesgo: tienen los mayores anchos dentro del espectro, una característica que no está relacionada con la longitud de sus secuencias genómicas (SGs). Adicionalmente, el cromosoma Y muestra el menor máximo, ligeramente desplazado a la derecha relativo a los otros máximos espectrales.

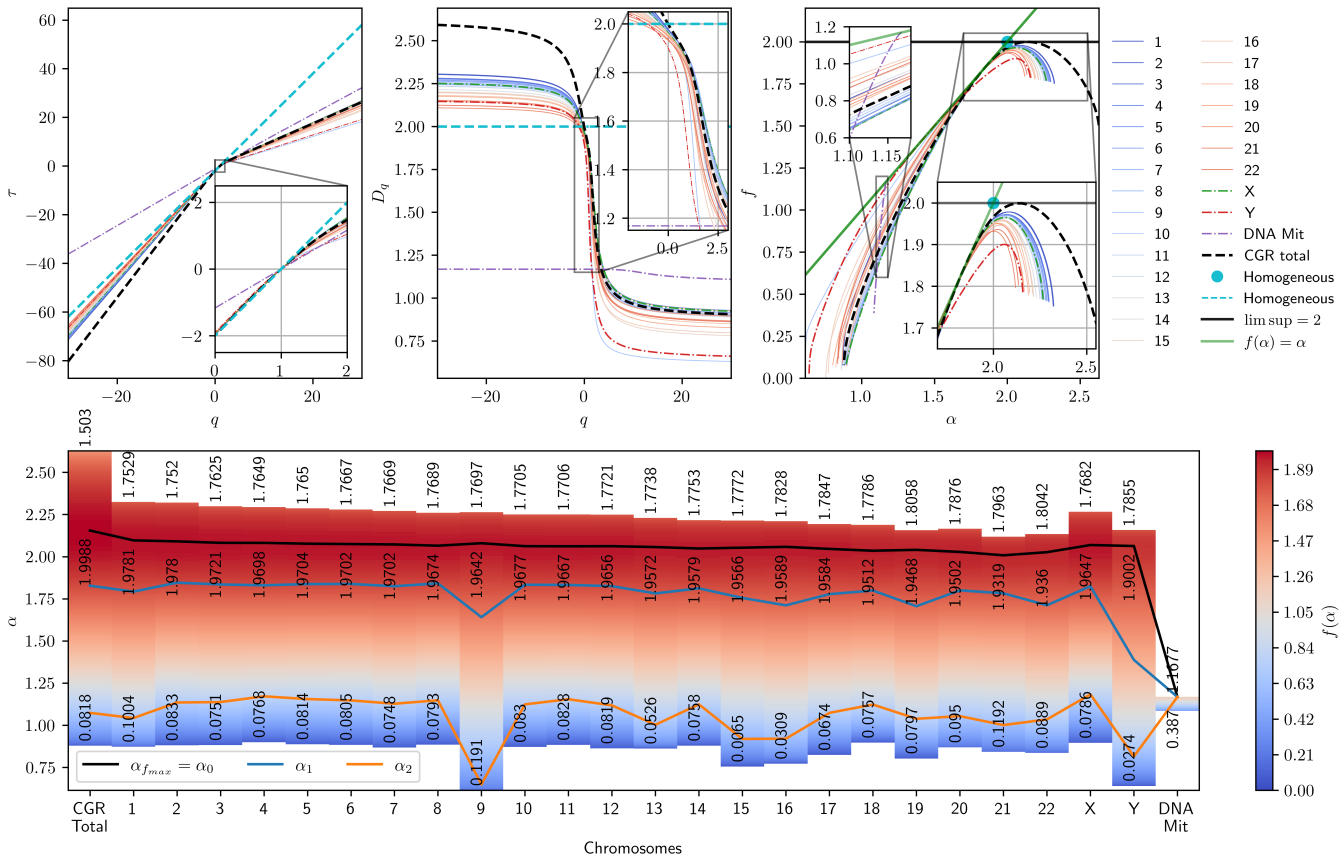
Las discrepancias para valores de  $q$  diferentes de cero son más evidentes; la parte inferior de la Figura 5.12 también muestra los valores correspondientes para la dimensión de información  $D_1$  y la dimensión de correlación  $D_2$ , siendo estas diferencias más pronunciadas en los cromosomas 9 e Y. Esto proporciona evidencia inicial de que estos cromosomas exhiben mayor heterogeneidad en secuencias de alta probabilidad. Además, la Figura 5.9 destaca que los valores máximos para estos cromosomas



**Figura 5.11:** Comparación unidimensional de cada cromosoma con el ensamblaje completo (CGR Total). Cada representación se genera alineando horizontalmente las filas de la distribución, como se ilustra en la Figura 5.10. Las densidades  $\rho$  y  $\rho_{\max}$  se muestran en escala logarítmica. Seguimos un orden ascendente en los CGRs, comenzando desde la fila correspondiente a la SG corta de 12 bases que va desde AAAAAAAAAAAAAA hasta TTTTTTTTTTTT, hasta la fila de CCCCCCCCCCCC a GGGGGGGGGGGG.

son más altos que en otras representaciones y se ubican en regiones menos frecuentes en comparación con las de otros cromosomas—regiones no predominantemente asociadas con repeticiones de los nucleótidos T y A. A pesar de esto, el cromosoma 9 muestra una menor diversidad de eventos de alta probabilidad, resultando en un rango de  $\alpha$  más estrecho en las áreas correspondientes. Estos hallazgos serán examinados con más detalle en la Sección 5.2.3. Cabe notar que las regiones pico están muy localizadas, ocurriendo en solo unos pocos lugares; en consecuencia, son menos representativas en comparación con las  $4^{12}$  ubicaciones posibles correspondientes a cada secuencia potencial, lo que hace más desafiante establecer diferencias multifractales en estas regiones.

Los resultados también indican que en todos los cromosomas la variación más amplia ocurre entre  $q = 1$  y  $q = 2$ , donde se observan los rangos de  $\alpha$  más amplios. Aunque esto no necesariamente indica la región de mayor diversidad genómica en la SG, sí implica un correspondiente aumento en



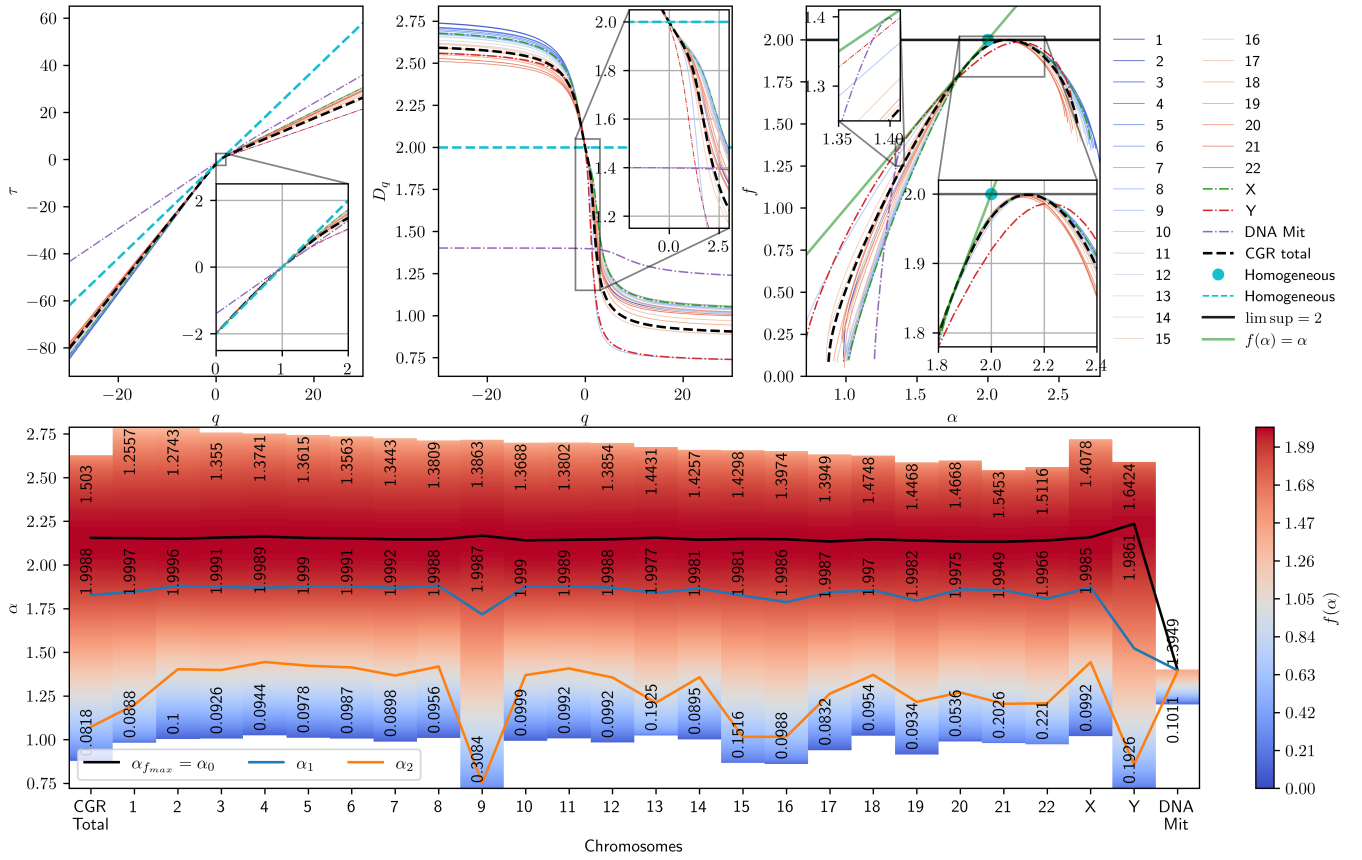
**Figura 5.12:** Análisis multifractal: el exponente  $\tau$  en función de  $q$ , la dimensión generalizada  $D_q$ , y el espectro multifractal  $f(\alpha)$ , donde  $\alpha$  representa el exponente de singularidad como se define en la Sección 5.0.4. Las gráficas para cada cromosoma se presentan y comparan con el ensamblaje completo. En cada secuencia, se usaron  $4^{12}$  cubrimientos, lo que corresponde a un tamaño de caja de  $r = 2^{-12}$ . También se incluyen valores correspondientes a la representación homogénea del cuadrado unitario. En la figura inferior,  $f(\alpha)$  para cada cromosoma se muestra usando un esquema de colores para representar las alturas correspondientes a cada valor de  $\alpha$ . Los valores numéricos corresponden a  $f(\alpha_{\min})$ ,  $f(\alpha_0)$ , y  $f(\alpha_{\max})$ . Adicionalmente, se proporcionan los valores correspondientes para las dimensiones fractal, de información y de correlación ( $\alpha_0$ ,  $\alpha_1$ , y  $\alpha_2$ ). El comportamiento entre diferentes cromosomas es generalmente similar, con la excepción de los cromosomas 9 e Y y el ADN mitocondrial.

la complejidad. Se encuentran variaciones significativas nuevamente en los cromosomas 9 e Y, donde las diferencias entre  $q = 0$  y  $q = 1$  son particularmente notables—especialmente para el cromosoma Y. En general, los cromosomas tienden a tener secuencias con solo pequeñas variaciones tanto en regiones de baja como de alta probabilidad, como lo evidencia el ancho estrecho  $w$  en esas regiones. En particular, el rango de  $\alpha$  para el cromosoma 9 en la región correspondiente a  $q > 2$  es el más estrecho, indicando una menor riqueza multifractal en esa área; en otras palabras, esta región es menos importante en términos de variación de secuencias.

En la Tabla 5.1, presentamos los valores promedio y las desviaciones estándar correspondientes para los cromosomas, excluyendo el ADN mitocondrial.

Respecto a las regiones de menor probabilidad—que cubren un área más grande dentro de la representación—se puede inferir que estas áreas están más cerca del soporte fractal y ocupan una región más amplia, ya que se distribuyen sobre un mayor número de secuencias de 12 nucleótidos que contienen el dinucleótido CG (una característica común a todos los cromosomas). Los cambios



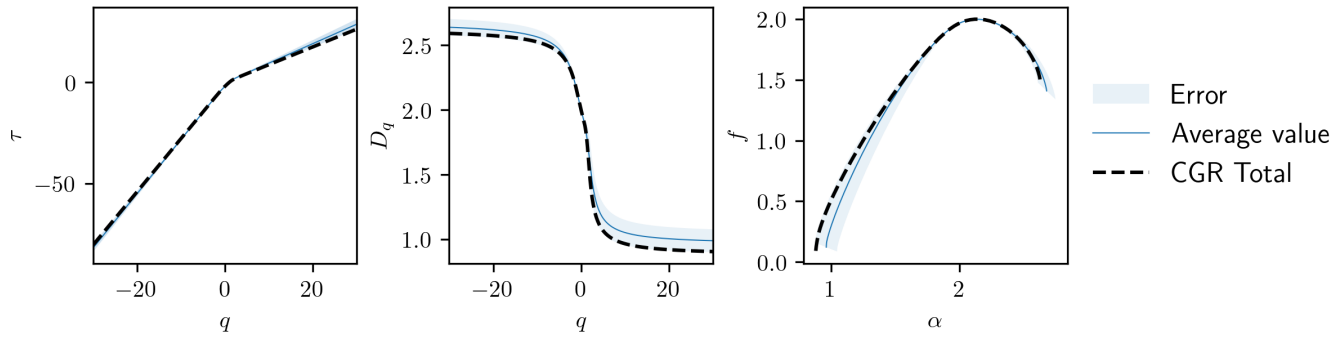


**Figura 5.13:** Análisis multifractal por cromosoma comparado con el ensamblaje completo (de la misma manera que se muestra en la Figura 5.12). El número de cubrimientos para los cromosomas se redujo a  $4^{10}$ , con un tamaño de caja de  $r = 2^{-10}$ , para mejorar el ajuste de los cromosomas con el ensamblaje para cada curva.

observados parecen estar más estrechamente asociados con la longitud de la SG de cada cromosoma. Existe una fuerte tendencia a que ocurran eventos de baja probabilidad; al ocupar una porción más grande del CGR, la suma de probabilidades en estas regiones excede la de los eventos locales. Adicionalmente, las curvas en los espectros exhiben alta asimetría relativa a sus valores máximos, con una desviación porcentual promedio ( $MDP = (\sum_{Chr} (|A_s|/f_{m\acute{a}x})_{Chr})/\#chr$ ) de  $(86.77 \pm 2.02) \%$ . Esto también puede interpretarse como que la mayoría de la población del CGR se encuentra fuera de las regiones de alta probabilidad local en todos los cromosomas.

El espectro promedio entre cromosomas y su desviación se obtuvo (ver Figura 5.14), para los resultados obtenidos en la Figura 5.13. El ensamblaje permanece dentro de la desviación a lo largo del espectro, con un mejor ajuste en el área alrededor del máximo.

Los máximos en los espectros se acercan a la dimensión del plano; sin embargo, dado que es un conjunto finito de puntos, en el límite infinitesimal, los espectros tienden a tener una dimensión de cero. Por lo tanto, el espectro multifractal sirve como una aproximación, asumiendo que la secuencia es considerablemente grande en relación con el número de cubrimientos y suficientemente pequeña para aproximarse al atractor de la distribución. No obstante, dado que el conjunto de cromosomas es significativamente más grande que las secuencias de cromosomas individuales, podríamos asumir



**Figura 5.14:** Comparación del exponente  $\tau$ ,  $D_q$ , y el espectro multifractal entre el ensamblaje CGR y el promedio cromosómico obtenido de los resultados en la Figura 5.13. Se usaron dos tipos de cubrimientos— $10^{12}$  para el CGR y  $10^{10}$  para los cromosomas—para compensar las diferencias en las longitudes de las secuencias. Las variaciones de escala son más pronunciadas para valores negativos de  $q$ , y el soporte fractal ( $D_0$ ) muestra un mejor acuerdo en promedio que cuando se usa un número mayor de cubrimientos.

**Tabla 5.1:** Valores promedio y desviaciones estándar de  $\tau(q)$ ,  $D_q$ ,  $\alpha_q$ , y  $f(\alpha_q)$  a través de los cromosomas para diferentes valores de  $q$ , derivados de los resultados mostrados en la Figura 5.12.

| $q$ | $\bar{\tau}(q)$  | $\bar{D}_q$     | $\bar{\alpha}_q$ | $\bar{f}(\alpha_q)$ |
|-----|------------------|-----------------|------------------|---------------------|
| -30 | $-68.78 \pm 1.7$ | $2.22 \pm 0.05$ | $2.23 \pm 0.06$  | $1.78 \pm 0.01$     |
| 0   | $-1.96 \pm 0.02$ | $1.96 \pm 0.02$ | $2.04 \pm 0.02$  | $1.96 \pm 0.02$     |
| 1   | $0.0 \pm 0.0$    | $1.81 \pm 0.08$ | $1.78 \pm 0.10$  | $1.78 \pm 0.10$     |
| 2   | $1.41 \pm 0.15$  | $1.41 \pm 0.15$ | $1.06 \pm 0.12$  | $0.70 \pm 0.11$     |
| 29  | $24.28 \pm 2.15$ | $0.87 \pm 0.08$ | $0.84 \pm 0.07$  | $0.08 \pm 0.03$     |

que su comportamiento promedio tendería a permanecer consistente, incluso si las longitudes de los cromosomas variaran.

El ADN mitocondrial es distinto de los otros espectros, exhibiendo un rango de  $\alpha$  muy estrecho de aproximadamente  $8 \times 10^{-2}$ . Esto se debe en parte a su longitud más corta pero principalmente porque su soporte difiere significativamente del de los cromosomas, como puede verse en las distribuciones. Sin embargo, su espectro proporciona nuevas perspectivas sobre su tendencia hacia dimensiones por debajo de valores de referencia como el del triángulo de Sierpinski, que tiene una dimensión fractal de aproximadamente 1.585; obtuvimos una dimensión fractal aproximada de 1.17. Este resultado probablemente se debe, en parte, a su sesgo hacia regiones correspondientes a las bases A y C, considerando que la base G representa solo alrededor del 13 % de la secuencia genómica—aproximadamente la mitad de la frecuencia de las otras bases. No se observaron cambios significativos para las dimensiones de información o correlación (ver Figura 5.12), lo que indica una mayor variación en secuencias de alta probabilidad correspondientes a  $q > 2$ ; los resultados se muestran en la Tabla 5.2. En este caso, las regiones asociadas con bajas probabilidades corresponden a secuencias cortas con un mayor contenido de G.

**Tabla 5.2:** Valores de  $\tau(q)$ ,  $D_q$ ,  $\alpha_q$ , y  $f(\alpha_q)$  para el ADN mitocondrial para diferentes valores de  $q$ , derivados de los resultados mostrados en la Figura 5.12.

| $q$ | $\tau(q)$ | $D_q$   | $\alpha_q$ | $f(\alpha_q)$ |
|-----|-----------|---------|------------|---------------|
| -30 | -36.20815 | 1.168   | 1.16802    | 1.16765       |
| 0   | -1.16784  | 1.16784 | 1.16789    | 1.16784       |
| 1   | 0.0       | 1.16777 | 1.16776    | 1.16776       |
| 2   | 1.16765   | 1.16765 | 1.1675     | 1.16734       |
| 29  | 31.06883  | 1.1096  | 1.08468    | 0.38699       |

En el apéndice A se presentan las representaciones obtenidas para distintos valores de  $M$  de cada cromosoma, tanto en dos como en tres dimensiones, con el propósito de complementar y ampliar los resultados expuestos en esta sección. Asimismo, se describen brevemente los criterios empleados para la selección del tamaño de las cubiertas utilizadas en el análisis multifractal y para la estimación de los espectros multifractales correspondientes.

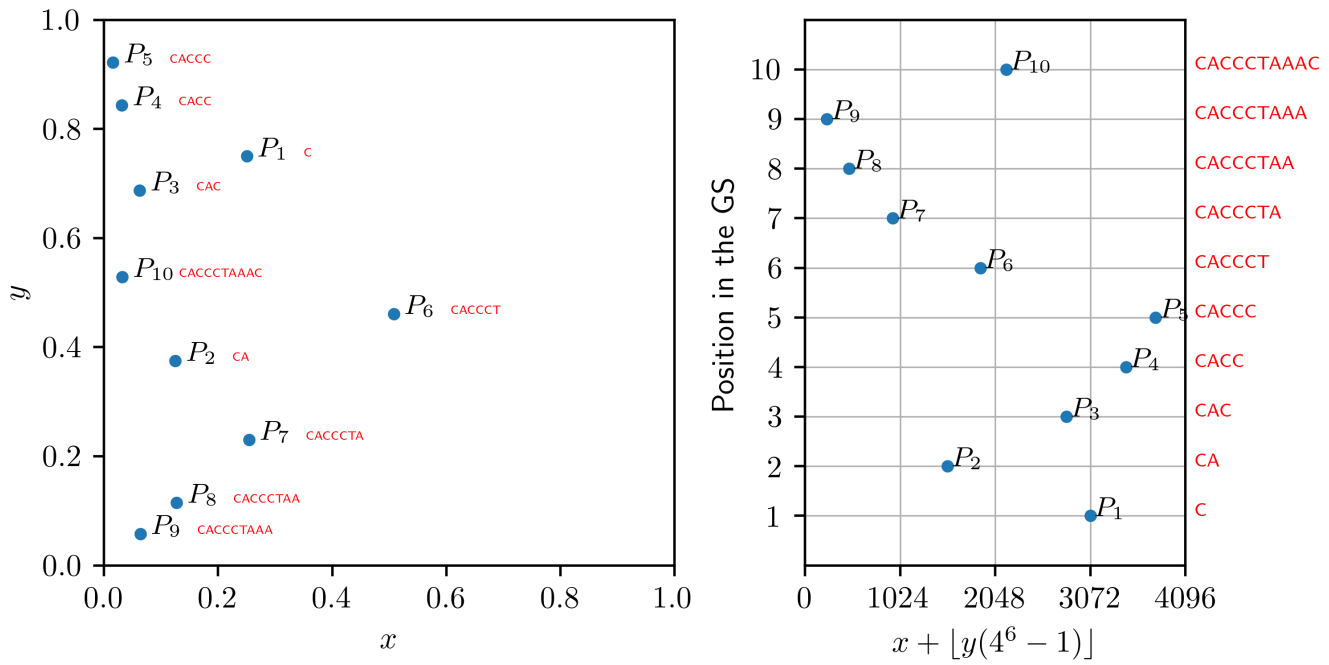
### 5.2.1. Citogenética del CGR

Dado que el CGR respeta el orden de la secuencia y la ubicación de sus puntos depende de estados anteriores, es posible extender el CGR unidimensional a lo largo de la secuencia para analizar cambios potenciales. La Figura 5.15 muestra la gráfica de la ubicación de cada base en la SG versus el CGR unidimensional obtenido mediante la transformación  $x + \lfloor y(4^6 - 1) \rfloor$ . Esta transformación organiza los puntos de manera similar a la Figura 5.10, extendiéndolos en un diseño unidimensional que sigue el orden de la secuencia genómica.

La Figura 5.16 muestra las gráficas correspondientes para la SG de cada cromosoma y la ubicación aproximada de la región del centrómero, tomada de las bases de datos en [8].

Este tipo de representación nos permite observar los patrones de bandas característicos de los cromosomas (Sección 5.0.1), lo que facilita, aproximadamente, la identificación de diferentes secciones de la SG dentro de los cromosomas o regiones enteras con concentraciones altas o bajas de ciertas bases. Por ejemplo, en el cromosoma 1, cerca del centrómero (marcado aproximadamente en rojo), podemos distinguir una región casi vacía de CG, junto con otras áreas más pequeñas con distribuciones sesgadas hacia las bases C y G. Esto podría indicar potenciales regiones codificantes y no codificantes. Sin embargo, la importancia de esta imagen desde una perspectiva multifractal radica en mostrar que la secuencia no tiene una estructura de soporte única a lo largo de la SG de los cromosomas.

Los cromosomas muestran grandes regiones más claras dentro de la secuencia—particularmente los cromosomas 1, 9, 15, 16 e Y—porque los puntos en la secuencia se acumulan alrededor de regiones



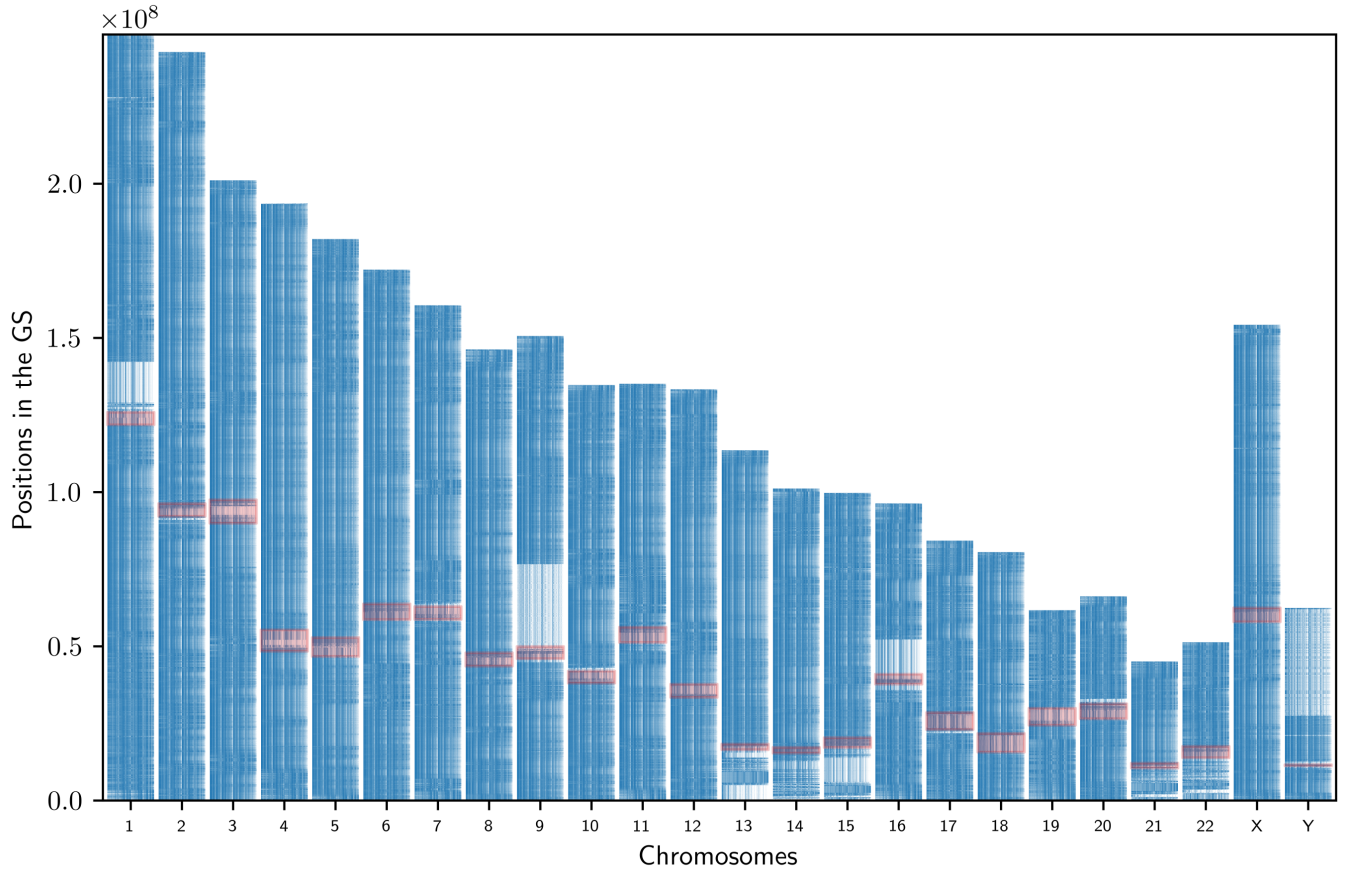
**Figura 5.15:** A la izquierda, la gráfica de los puntos iniciales de la SG para el cromosoma 1 usando CGR, y a la derecha, la gráfica de la ubicación de la base de la SG a lo largo del eje y comparada con el CGR unidimensional obtenido usando la transformación  $x + \lfloor y(4^6 - 1) \rfloor$  a lo largo del eje x.

más pequeñas en el CGR, a pesar de que la densidad de puntos en cada segmento permanece igual. En términos de secuencias genómicas, esto indica que estas áreas contienen una mayor presencia de secuencias repetitivas. Por el contrario, las regiones más oscuras muestran una mayor diversidad en los tipos de secuencias, con una distribución más homogénea de los cuatro nucleótidos organizados en un patrón heterogéneo que cubre un área mayor. En algunas bandas, las variaciones en la concentración en un lado u otro pueden indicar la ausencia de una combinación específica de nucleótidos.

También notamos que la mayoría de las áreas más claras se ubican en regiones pericentroméricas (cerca del centrómero), lo que puede ayudar a distinguir tipos de cromosomas según la posición del centrómero y su propensión a secuencias repetitivas. En consecuencia, se puede inferir que una gran parte de los picos de alta frecuencia está asociada con estas bandas más claras y repetitivas.

Aunque la resolución de la Figura 5.16 no permite identificar variaciones en regiones pequeñas, es importante notar que las bandas también pueden observarse a escalas mucho menores, demostrando aún más la autosimilitud y multifractalidad asociadas con bandas específicas. Para los propósitos de este estudio, solo pretendemos resaltar estas características cualitativas de los CGRs para secuencias genómicas.

El análisis base por base es común en estudios genómicos, ya que permite a los investigadores anotar cada sección cromosómica en detalle. Esta representación, combinada con una medida dimensional, puede ayudar en la ubicación y caracterización de regiones de interés.



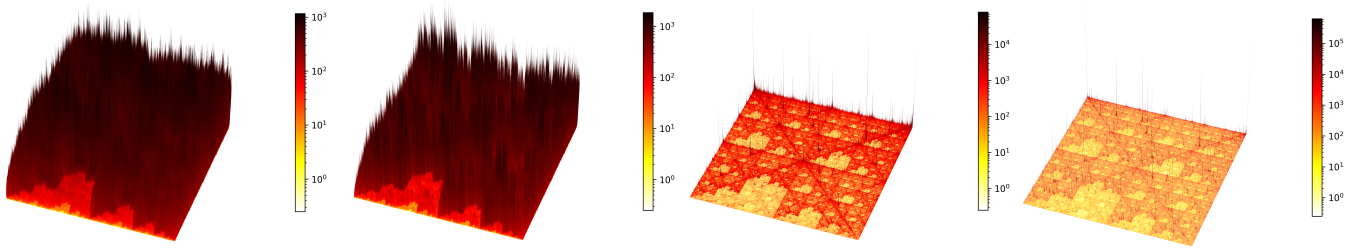
**Figura 5.16:** SG completa para cada cromosoma con soporte geométrico en el CGR. Los puntos de la representación del juego del caos se muestran unidimensionalmente usando la transformación  $x + \lfloor y(4^6 - 1) \rfloor$  para cada punto, mientras se preserva su orden secuencial en la SG, donde el eje horizontal para cada cromosoma varía desde 0 hasta  $4^6$ . Los marcadores rojos indican las ubicaciones aproximadas de los centrómeros.

### 5.2.2. Métodos de CM y BGR

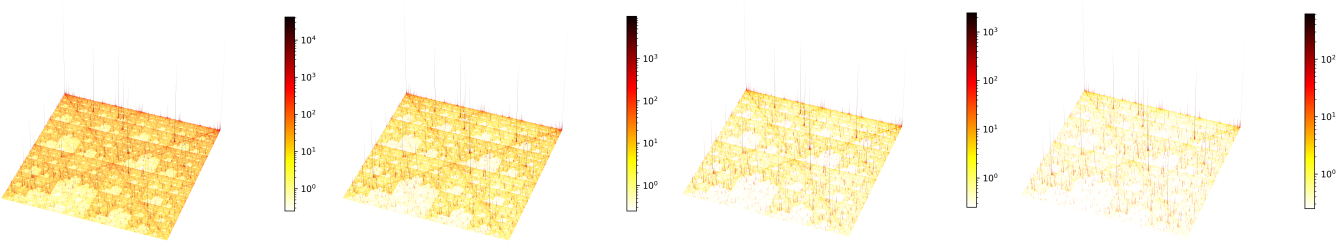
Dado que el ensamblaje cromosómico exhibe un comportamiento cercano al promedio cromosómico, podemos usarlo como referencia para evaluar los métodos de compresión, codificación y decodificación de CM y BGR 5.0.2. Las Figuras 5.17 y 5.18 muestran distribuciones de puntos para secuencias generadas por estos métodos, usando diferentes parámetros de longitud. Para CM, se consideraron secuencias de varias longitudes ( $n$ ), creando matrices de probabilidad para dinucleótidos, trinucleótidos, hexanucleótidos y dodecanucleótidos. Cada matriz se usó para generar secuencias y CGRs que coincidieran con la longitud de la SG del ensamblaje, es decir, 3,117,275,501 bases. Observamos diferencias notables en regiones fuera de las áreas CG, especialmente para cadenas más cortas; sin embargo, el ajuste mejora para secuencias de 6 y 12 bases.

El comportamiento es diferente para el método BGR, donde, dependiendo de la longitud de las particiones  $m$ , el número de puntos generados a partir de la SG es mucho menor que el del ensamblaje original. Esta reducción impacta principalmente las áreas menos representadas mientras preserva la estructura en regiones más densas.

Los sesgos de distribución observados con ambos métodos son evidentes en sus espectros multi-



**Figura 5.17:** Distribuciones de puntos generadas a partir de cadenas de Markov, con cuatro longitudes de nucleótidos diferentes ( $n = 2, 3, 6, 12$ ) mostradas de izquierda a derecha. En cada caso, se calcularon las matrices de Markov correspondientes a partir de la secuencia original y luego se usaron para generar una nueva secuencia aleatoria de la misma longitud. Aunque los resultados varían considerablemente con diferentes valores de  $n$ , exhiben una tendencia hacia la convergencia. Cada representación empleó  $4^{12}$  cubrimientos.



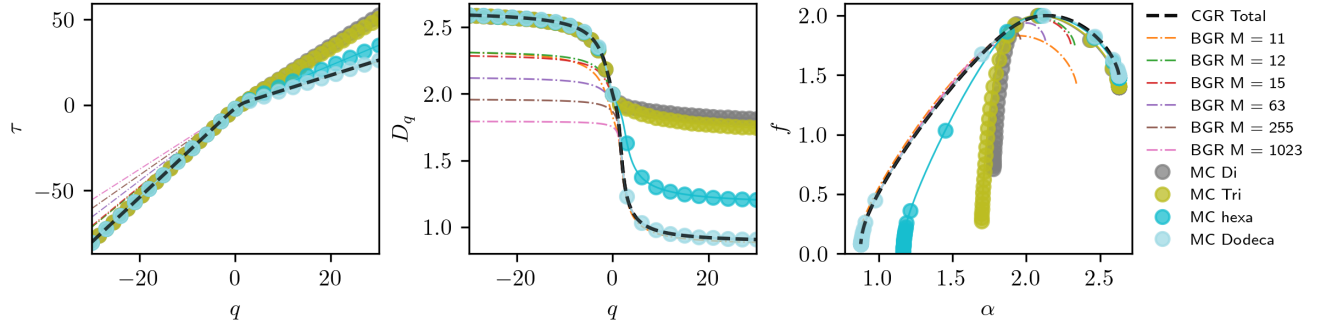
**Figura 5.18:** Distribuciones de puntos generadas a partir de BGR para  $M = \{15, 63, 255, 1023\}$  mostradas de izquierda a derecha. En cada representación, el conteo se realizó usando  $4^{12}$  cubrimientos. La longitud de la secuencia afecta el número de puntos que forman el BGR, impactando principalmente áreas de baja densidad.

fractales comparados con el CGR completo. La Figura 5.19 muestra que, en áreas de baja incidencia ( $q < 0$ ) y regiones CG del CGR (ver Figura 5.3), el método BGR pierde precisión a medida que  $M$  aumenta, impactando zonas de codificación con tasas de compresión más altas. Sin embargo, el ajuste en regiones no codificantes de alta incidencia permanece casi perfecto. Es importante destacar que BGR permite la decodificación completa de la secuencia, manteniendo características de distribución en zonas no codificantes.

En CM, las secuencias tienden a ajustarse mejor en áreas con menor densidad de población porque su longitud coincide con la del ensamblaje. Sin embargo, las secuencias más cortas pueden afectar negativamente a regiones con baja incidencia. La precisión mejora significativamente con particiones más largas, para valores mayores de  $n$ -nucleótido, logrando un ajuste casi perfecto con cadenas de 12 bases. Esto sugiere que la longitud óptima de secuencia para CM debería corresponder al número de particiones en la distribución del CGR ( $4^{12}$  cajas). Aunque CM se alinea bien con el espectro multifractal del ensamblaje, no es un método de codificación ni de decodificación, ya que cada valor se genera aleatoriamente y no refleja los patrones de bandas cromosómicas, como se muestra en la Figura 5.16. No obstante, el ajuste multifractal es notable, con un error promedio de aproximadamente 2% relativo al ensamblaje en secuencias de 12 bases.

Los ajustes observados en la Figura 5.19 ayudan a determinar qué método usar dependiendo de si se desea estudiar regiones de alta o baja probabilidad. Así, para regiones de alta probabilidad (lado





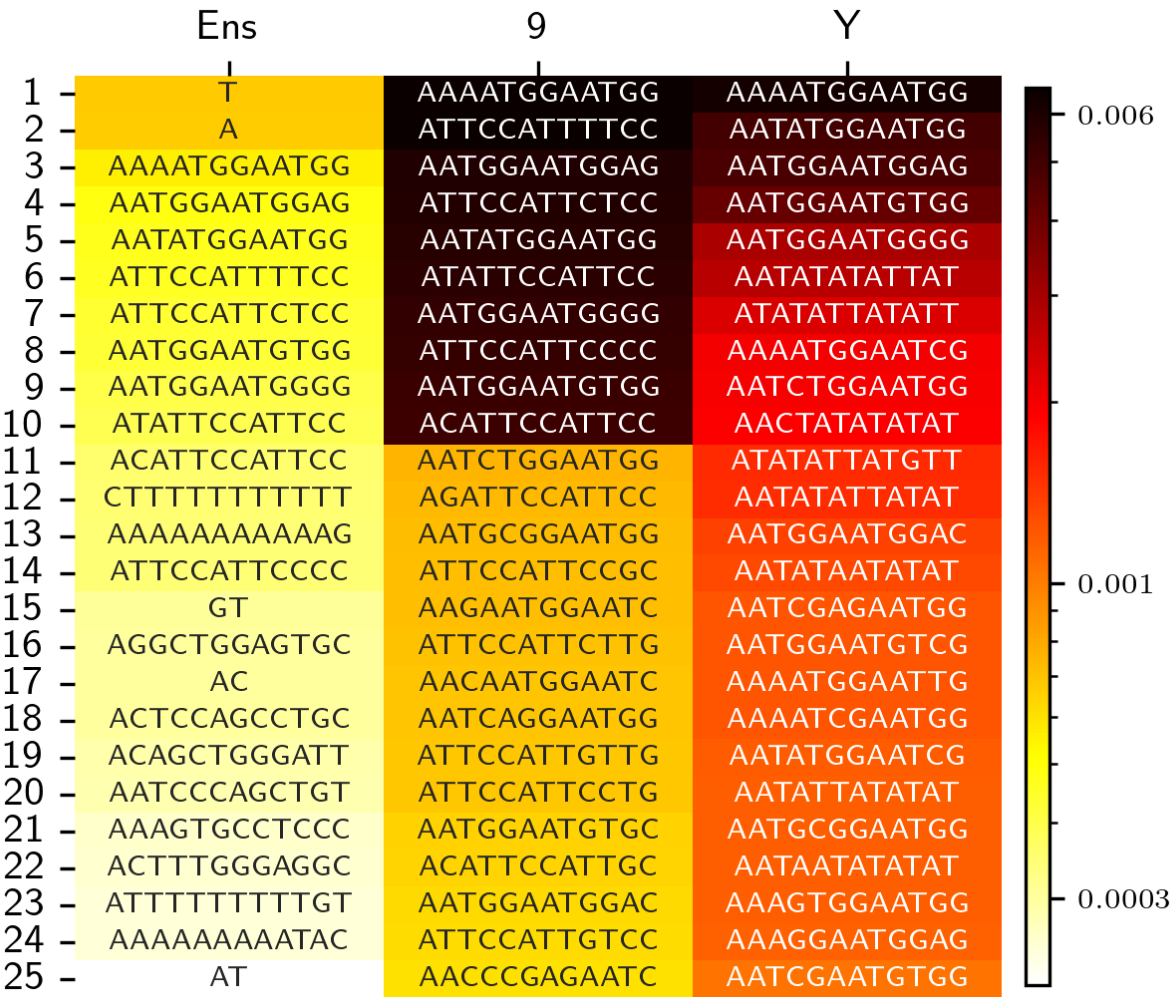
**Figura 5.19:** Análisis multifractal de los métodos BGR y CM comparado con los resultados del CGR completo. En estas gráficas, se muestran los valores de  $\tau(q)$ ,  $D_q$ , y  $f(\alpha)$  para el CGR del ensamblaje cromosómico completo, junto con aquellos obtenidos por los métodos CM y BGR. CM proporciona un mejor ajuste en el lado derecho del espectro  $f(\alpha)$  y en el máximo del CGR; sin embargo, en el lado izquierdo, se ajusta con precisión solo para las secuencias de dodecanucleótidos. Por el contrario, BGR generalmente se ajusta bien en el lado izquierdo a altas frecuencias para cualquier valor de  $M$ .

izquierdo del espectro), es más conveniente usar BGR con cualquier valor de  $M$ , mientras que para regiones de baja probabilidad (lado derecho del espectro), CM siempre funciona mejor. En general, estos métodos pueden ser complementarios y facilitar el análisis de secuencias genómicas (SGs). Sin embargo, estos resultados están limitados a secuencias largas, ya que el estudio se realizó en comparación con el espectro del ensamblaje completo. En consecuencia, el comportamiento puede variar para longitudes más cortas o secciones específicas del genoma, teniendo en cuenta el comportamiento en bandas observado en la SG de cada cromosoma.

### 5.2.3. Ejemplo de Aplicación en los Cromosomas 9 e Y

En esta sección, discutimos un enfoque para aplicar los resultados obtenidos mediante CGR. Para este ejemplo, usamos los cromosomas 9 e Y porque exhiben el rango más amplio en el exponente de Hölder. Sin embargo, nos enfocamos solo en una característica específica del espectro: las regiones de mayor singularidad en la distribución del CGR. Estas regiones corresponden a los picos más altos, que son más pronunciados en estos cromosomas que en otros y se ubican fuera de las regiones asociadas con repeticiones de bases A y T.

Basándonos en la distribución del CGR, se identificaron secuencias altamente frecuentes para el ensamblaje completo así como para los cromosomas 9 e Y, como se muestra en la Figura 5.20. Las secuencias de 12 nucleótidos identificadas incluyen todas las rotaciones cíclicas posibles (por ejemplo, ACGTFGCAACGTFGCA  $\rightarrow$  AACGTFGCAACGTFGC  $\rightarrow$  CAACGTFGCAACGTFG  $\rightarrow$  ...); por lo tanto, solo se muestra la primera secuencia en orden alfabético. Este enfoque ayuda a descartar la posibilidad de que las singularidades correspondan meramente a rotaciones de otras secuencias, ya que las repeticiones perfectas inherentemente incluyen todas las rotaciones cuando se leen base por base. Considerar estas rotaciones cíclicas suma hasta 11 secuencias adicionales ubicadas en diferentes



**Figura 5.20:** Las 25 secuencias con las frecuencias más altas para la SG de los cromosomas 9 e Y, comparadas con las del ensamblaje completo. Cada secuencia corresponde a la primera rotación cíclica en orden alfabético. Los colores indican el valor de probabilidad (frecuencia relativa a la longitud total de la SG del ensamblaje o las longitudes cromosómicas correspondientes) asociado con cada secuencia. Para secuencias repetitivas, se asigna el patrón más corto (por ejemplo, “T” corresponde a “TTTTTTTTTTTTT”, o “GT” a “GTGTGTGTGTGT”).

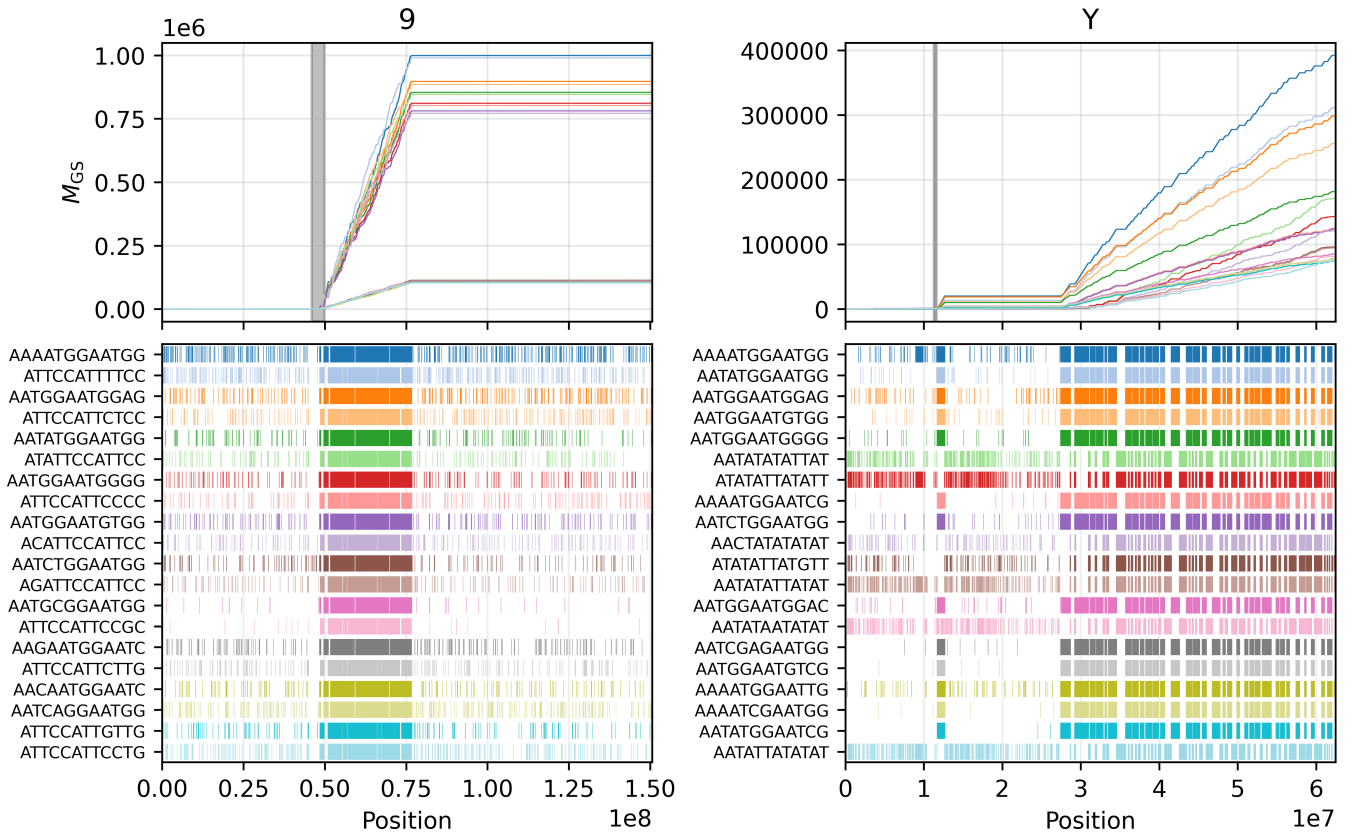
regiones de la distribución. Así, si se considerara el CGR completo, el número de secuencias distintas se reduciría a menos de  $4^{12}$ , dado que la mayoría de las secuencias (por ejemplo, aquellas compuestas de una sola base repetida) carecen de variantes cíclicas o tienen menos de ellas.

Algunas de las secuencias en la Figura 5.20 corresponden a polimorfismos—variaciones en una o pocas bases—típicas de regiones altamente repetitivas o heterocromatina. En genómica, los polimorfismos se refieren a la presencia de dos o más formas variantes de una secuencia de ADN específica que pueden ocurrir entre diferentes individuos o poblaciones. El tipo más común de polimorfismo implica variación en un solo nucleótido (conocidos como polimorfismos de nucleótido único, o SNPs). Otros polimorfismos pueden ser mucho más grandes, abarcando segmentos de ADN más largos. Estas variaciones son significativas porque introducen diversidad en el genoma, haciendo que su caracterización sea particularmente útil. Los polimorfismos también explican por qué el ancho del espectro multifractal asociado con regiones de alta probabilidad en el cromosoma 9 es más estrecho. Estas



variaciones localizadas, que afectan solo porciones pequeñas de la secuencia, conducen a una mayor homogeneidad en estas regiones en comparación con otras secciones de la secuencia genómica (SG). En otras palabras, los polimorfismos introducen variaciones menores en términos multifractales.

Además de las secuencias de 12 bases, varias secuencias más cortas se observaron recurrentemente en los datos. Por ejemplo, AATGGAATG aparece al menos ocho veces en el cromosoma 9, mientras que repeticiones simples como TAT o eventos disruptivos de inserción/delección (por ejemplo, ATTCC → ATCG) se encuentran en el cromosoma Y. Estas observaciones se alinean con la arquitectura genómica descrita en la literatura [19] para ADN heterocromático y repeticiones de transposones, donde la heterocromatina constitutiva en centrómeros y otras regiones consiste principalmente en arreglos muy largos de repeticiones en tandem de alto número de copias, conocidas como ADN satélite.

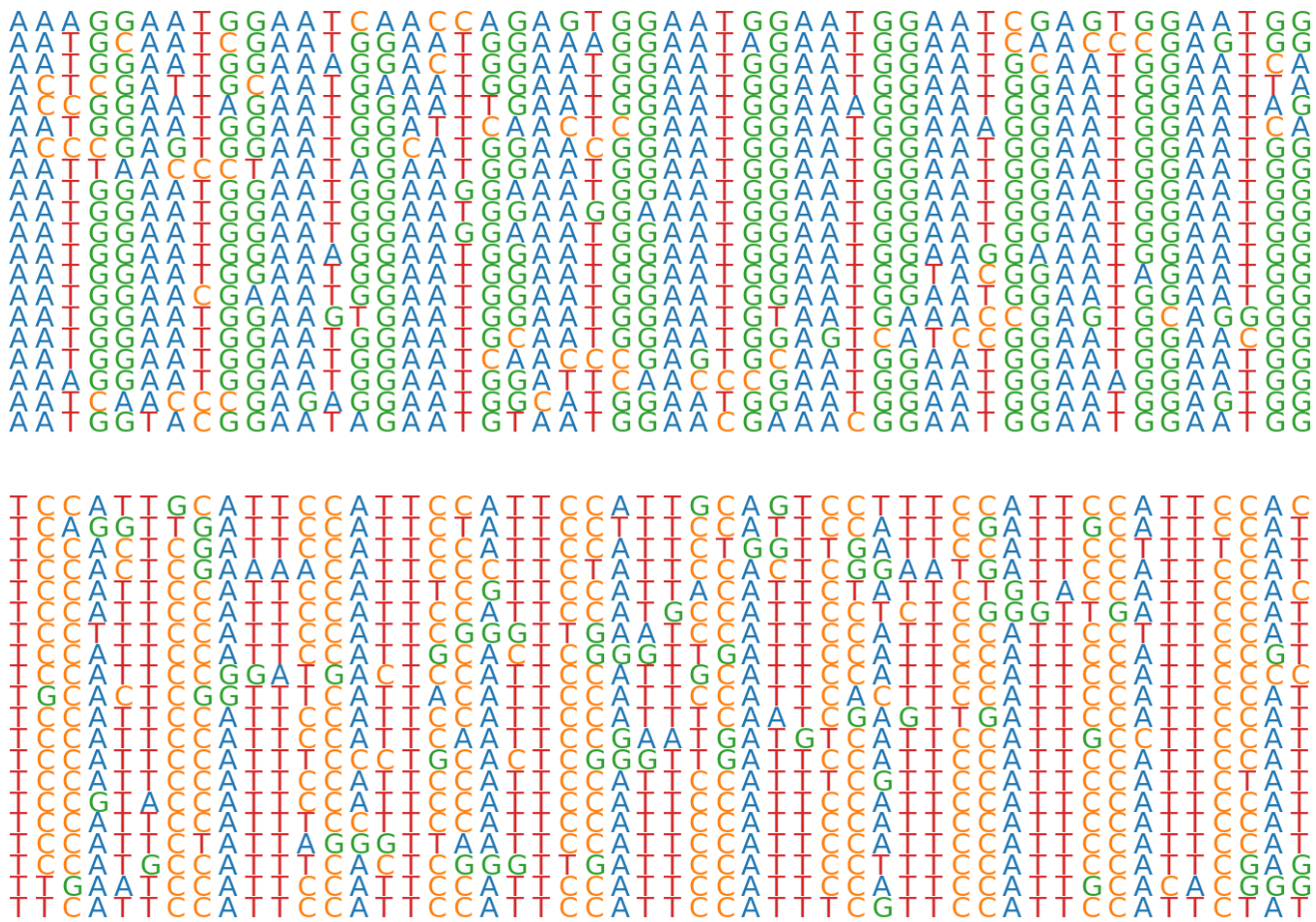


**Figura 5.21:** Se muestra la secuencia de mayor frecuencia en los cromosomas 9 e Y. El panel inferior resalta las posiciones en la SG con la mayor incidencia de esta secuencia, mientras que el panel superior muestra las frecuencias acumulativas  $M_{GS}$  correspondientes, enfatizando las regiones de mayor densidad donde se observa la tasa de crecimiento máxima. La banda gris marca la ubicación aproximada del centrómero.

Usando las secuencias obtenidas, mapeamos sus posiciones a lo largo de las SGs. En la Figura 5.21, se marcan las ubicaciones de las secuencias asociadas con los picos de distribución, identificando regiones con la mayor densidad de secuencias. El panel superior muestra las frecuencias acumulativas para cada secuencia. Los resultados revelan tanto regiones con una alta concentración de estas secuencias como áreas donde están completamente ausentes. Esto sugiere que las regiones identi-

cadascorresponden a secciones repetitivas del genoma, reflejando sus características estructurales y biológicas asociadas. Estas regiones de alta densidad también coinciden con las áreas más claras catalogadas en la Figura 5.16 para los dos cromosomas correspondientes.

Analizar las frecuencias de secuencia a lo largo de la SG nos permite identificar regiones con la mayor concentración de la secuencia objetivo. La Figura 5.22 presenta dos regiones del cromosoma 9 con la mayor ocurrencia de secuencias de alta frecuencia. Estos segmentos exhiben periodicidad en las secuencias AATGG y ATTCC (satélites 2 y 3), aunque interrumpidas por pequeñas “imperfecciones” o polimorfismos. En consecuencia, los picos en la secuencia caracterizan estos polimorfismos según su frecuencia o probabilidad de aparición. Dado que podemos distinguir dos tipos diferentes de secuencia en el cromosoma 9—correspondientes a repeticiones AATGG y ATTCC en las mismas regiones—estas secuencias también deberían estar intercaladas dentro de esta región.



**Figura 5.22:** Se muestran dos segmentos de la secuencia genómica en el cromosoma 9: el segmento superior abarca posiciones de nucleótidos 68,154,764 a 68,155,764, y el segmento inferior abarca posiciones 55,833,213 a 55,834,213. Estas regiones corresponden a las mayores ocurrencias de las secuencias AATGGAATGGAA y TTCCATTCCATT, respectivamente, incluyendo sus rotaciones cíclicas.

En el cromosoma Y, observamos secuencias que también se encuentran en el cromosoma 9, generalmente representando polimorfismos del motivo AAAATGGGAATGG. Sin embargo, están presentes formas distintas de repeticiones AT, que comúnmente se asocian con las subfamilias de satélite 1

y DYZ2. DYZ2 se usa típicamente para detectar anomalías clínicas así como variaciones dentro de la población normal. Se observan dos regiones separadas—una cerca del centrómero y otra al final del brazo largo—pero estas regiones están altamente fragmentadas, con áreas que exhiben poca o ninguna población de secuencias.

Este análisis puede generalizarse a todos los cromosomas. Sin embargo, esta es solo una forma en que CGR, CM o BGR pueden usarse para estudiar SGs. Muchos otros enfoques de análisis multifractal podrían emplearse en estos estudios. Por ejemplo, se podrían analizar regiones basadas en su medida de probabilidad y asociarlas con bandas y regiones de la SG. Otra aplicación potencial es la identificación y caracterización de polimorfismos entre individuos o especies.



# Conclusiones y perspectivas

Este trabajo de investigación ha desarrollado y aplicado metodologías para el análisis estructural de secuencias genómicas mediante la Representación Binaria Genómica (BGR) y el análisis multifractal. Los resultados obtenidos proporcionan herramientas y perspectivas para caracterizar la organización de la información genética, superando algunas limitaciones de los enfoques tradicionales y estableciendo bases para futuras investigaciones en genómica computacional.

La Representación Binaria Genómica constituye una adaptación de la CGR que aborda problemas de precisión numérica y escalabilidad identificados en enfoques previos. Al expresar los nucleótidos como vectores binarios  $A = (0, 0)$ ,  $C = (0, 1)$ ,  $G = (1, 1)$ ,  $T = (1, 0)$ , se demostró que el CGR puede interpretarse como la trayectoria de un caminante determinista donde cada punto contiene información sobre la secuencia previa. Esto permite un tratamiento matemático más riguroso de las representaciones genómicas y facilita el análisis computacional de secuencias de gran longitud.

La introducción del parámetro  $M$  en el esquema de Caminante Determinista Compactado representa una contribución práctica significativa, ya que permite controlar de manera sistemática la resolución de las representaciones según los requerimientos específicos de cada análisis. Este control sobre el nivel de detalle posibilita adaptar el método a diferentes tipos de estudios, desde análisis globales del genoma completo hasta investigaciones focalizadas en regiones específicas.

La equivalencia establecida entre los esquemas de Jeffrey y Yin mediante el formalismo BGR permite el desarrollo de esquemas de codificación mediante vectores enteros que reducen sustancialmente los requerimientos de almacenamiento mientras mantienen la integridad de la información estructural. Esta capacidad de codificación compacta, combinada con la propiedad de que el último punto de la trayectoria contiene información completa sobre la secuencia, facilita la reconstrucción precisa de secuencias a partir de representaciones minimalistas, lo que tiene implicaciones importantes para el almacenamiento y procesamiento eficiente de datos genómicos.

El análisis multifractal aplicado al genoma humano T2T-CHM13v2.0 revela patrones estructurales que presentan heterogeneidades consistentes a múltiples escalas. La estructura multifractal se mantiene de manera estable a través de la secuencia completa, observándose patrones similares tanto en el ensamblaje global como en los cromosomas individuales, lo que sugiere la existencia de principios organizativos comunes en la arquitectura genómica.

Los espectros multifractales calculados mediante técnicas de box-counting sistemáticas indican que cada cromosoma posee características multifractales distintivas, con variaciones que reflejan diferencias en su composición y organización estructural. Estos resultados proporcionan una base cuantitativa para la caracterización comparativa de cromosomas y para el estudio de las relaciones entre estructura genómica y función biológica.

Los cromosomas 9 e Y presentan comportamientos diferenciados en comparación con otros cromosomas, mostrando rangos más amplios de exponentes de Hölder y características específicas en sus espectros multifractales. Estas diferencias estructurales merecen mayor investigación para comprender su relación con las funciones biológicas particulares de estos cromosomas.

El CGR demuestra capacidad consistente para distinguir entre diferentes tipos de regiones genómicas, incluyendo la separación entre regiones codificantes y no codificantes, la identificación de islas CpG/GpC y la detección de secuencias repetitivas en centrómeros y regiones heterocromáticas. Estas capacidades sustentan la utilidad del análisis multifractal como herramienta complementaria para la caracterización estructural del genoma.

La comparación sistemática entre representaciones BGR y de Cadenas de Markov revela complementariedades analíticas importantes para diferentes tipos de estudios genómicos. La BGR muestra mayor eficacia para capturar componentes de alta frecuencia asociados con regiones repetitivas y secuencias conservadas, mientras que las Cadenas de Markov se alinean mejor con regiones de baja frecuencia ricas en dinucleótidos CG y áreas de mayor variabilidad.

El error promedio de 2 % obtenido para cadenas de doce nucleótidos usando el enfoque de Cadenas de Markov indica que este método puede simular aspectos significativos de la estructura multifractal del genoma. Sin embargo, la naturaleza estocástica inherente a las Cadenas de Markov limita su capacidad para replicar los patrones de bandas fijos observados en los cromosomas, lo que sugiere la conveniencia de usar este método principalmente para estudios donde esta limitación no sea crítica.

El análisis detallado de los cromosomas 9 e Y permitió identificar secuencias repetitivas específicas como AATGGAATGGAA y TTCCATTCCATT en regiones heterocromáticas, así como localizar con precisión polimorfismos y variaciones estructurales. Los patrones geométricos observados en las representaciones BGR y CGR muestran correlaciones consistentes con características funcionales del genoma, lo que respalda la validez biológica de estos enfoques.

Las regiones repetitivas identificadas no corresponden principalmente a secuencias mononucleotídicas simples, sino a motivos más complejos como los pentámeros ATTCC/GGAAT que interactúan con proteínas de heterocromatina y participan en mecanismos específicos de empaquetamiento cromosómico. Estos hallazgos contribuyen a comprender mejor la organización estructural del genoma y sus bases moleculares.

Desde la perspectiva computacional, la BGR representa una alternativa más eficiente al CGR tradicional, reduciendo significativamente los requerimientos de almacenamiento mediante esquemas

de codificación binaria y permitiendo análisis a múltiples escalas mediante la variación controlada del parámetro  $M$ . La posibilidad de realizar estudios comprehensivos a diferentes escalas sin emplear ventanas deslizantes computacionalmente costosas, manteniendo simultáneamente la autosimilitud y la información esencial, ofrece ventajas prácticas importantes para el análisis de genomas extensos.

Sin embargo, persisten desafíos computacionales para valores grandes de  $M$ , particularmente en lo relacionado con los requerimientos de almacenamiento y procesamiento. Estos desafíos sugieren la necesidad de desarrollar estrategias de compresión más avanzadas basadas en las propiedades fractales identificadas y en principios de teoría de la información.

Los métodos desarrollados en este trabajo son aplicables a genomas de diferente longitud y complejidad, proporcionando herramientas versátiles para diversos tipos de investigaciones genómicas. Entre las aplicaciones potenciales se incluyen análisis comparativos entre especies, identificación precisa de variantes estructurales en contextos de genómica médica y estudios evolutivos de arquitecturas genómicas a gran escala.

La capacidad demostrada para identificar regiones de interés biológico mediante características puramente geométricas sugiere posibles aplicaciones en el diagnóstico de anomalías cromosómicas y en el desarrollo de marcadores estructurales para condiciones genéticas. Estas aplicaciones podrían complementar los métodos tradicionales basados en secuenciación, particularmente en contextos donde los enfoques geométricos ofrezcan ventajas específicas.

Como perspectivas futuras inmediatas, se considera prometedora la integración de métodos BGR y de Cadenas de Markov para capturar simultáneamente características estructurales fijas y variaciones aleatorias. Esta integración podría permitir un análisis más comprehensivo de la organización genómica, combinando las fortalezas de ambos enfoques.

El análisis detallado de fluctuaciones locales a lo largo de cada cromosoma representa otra dirección importante para futuras investigaciones, particularmente para correlacionar variaciones estructurales específicas con funciones biológicas concretas. Asimismo, el estudio especializado del ADN mitocondrial, dada su estructura única y comportamiento multifractal diferenciado, merece atención particular.

El desarrollo de técnicas que establezcan correlaciones cuantitativas entre características multifractales específicas y funciones biológicas representa una dirección de investigación particularmente relevante. Estos desarrollos podrían contribuir significativamente a comprender mejor las relaciones entre estructura genómica y función biológica.

La demostración de que es posible identificar localizaciones de interés biológico a partir de características geométricas sin requerir búsquedas exhaustivas secuencia por secuencia establece un enfoque analítico alternativo que puede complementar metodologías tradicionales de secuenciación. Esta capacidad adquiere especial relevancia en el contexto de genomas con regiones repetitivas extensas, donde los métodos convencionales pueden enfrentar limitaciones prácticas.

En resumen, este trabajo ha desarrollado un marco metodológico que combina precisión matemática y escalabilidad computacional para el análisis estructural de secuencias genómicas. La integración de BGR y análisis multifractal contribuye a la comprensión de la organización genómica y proporciona herramientas prácticas para explorar la estructura de la información genética en diferentes escalas. Los resultados obtenidos indican que patrones geométricos cuantificables están presentes en las secuencias genómicas, lo que sugiere que la geometría fractal y el análisis multifractal pueden ofrecer perspectivas valiosas para describir la organización de sistemas biológicos complejos.

Las metodologías desarrolladas y los resultados obtenidos constituyen una base sólida para futuras investigaciones en genómica estructural y representan una contribución al desarrollo de herramientas computacionales para el análisis de datos genómicos a gran escala. El trabajo abre varias líneas de investigación promisorias que podrían conducir a avances significativos en nuestra comprensión de la organización fundamental de la información genética.



# Bibliografía

- [1] B. B. Mandelbrot, *The Fractal Geometry of Nature*. New York: W. H. Freeman and Company, 1982.
- [2] M. F. Barnsley, *Fractals Everywhere*. San Diego, USA: Morgan Kaufmann, 1993.
- [3] M. Schroeder, *Fractals, Chaos, Power Laws*. New York: W H Freeman and Company, 1991.
- [4] N. Goldman, “Nucleotide, dinucleotide and trinucleotide frequencies explain patterns observed in chaos game representations of dna sequences,” *Nucleic Acids Research*, vol. 21, no. 10, pp. 2487–2491, 1993.
- [5] K. R. SREENIVASAN and C. MENEVEAU, “The fractal facets of turbulence,” *Journal of Fluid Mechanics*, vol. 173, pp. 357–386, 1986.
- [6] H. Jeffrey, “Chaos game representation of gene structure,” *Nucleic Acids Research*, vol. 18, pp. 2163–2170, 04 1990.
- [7] Y. A. Álvarez Ballesteros, *Sistema de funciones iteradas, fractales y multifractales*. CDMX: Universidad Autónoma Metropolitana, 2021.
- [8] T2T, “Genome assembly T2T-CHM13v2.0.” [https://www.ncbi.nlm.nih.gov/datasets/genome/GCF\\_009914755.1/](https://www.ncbi.nlm.nih.gov/datasets/genome/GCF_009914755.1/), 2022.
- [9] T. C. Halsey, M. H. Jensen, L. P. Kadanoff, I. Procaccia, and B. I. Shraiman, “Fractal measures and their singularities: The characterization of strange sets,” *Phys. Rev. A and Nuclear Physics B - Proceedings Supplements*, vol. 33, 2, pp. 1141–1151, 501–511, Feb 1986, 1987. <https://link.aps.org/doi/10.1103/PhysRevA.33.1141>, <https://www.sciencedirect.com/science/article/pii/0920563287900363>.
- [10] J. Feder, *Fractals*. New York: Springer Science+Business Media, LLC, 1988.
- [11] A. Chhabra and R. V. Jensen, “Direct determination of the singularity spectrum,” *Phys. Rev. Lett.*, vol. 62, pp. 1327–1330, Mar 1989.

- [12] A. N. Kolmogorov, “A refinement of previous hypotheses concerning the local structure of turbulence in a viscous incompressible fluid at high reynolds number,” *Journal of Fluid Mechanics*, vol. 13, no. 1, p. 82–85, 1962.
- [13] B. B. Mandelbrot, “Possible refinement of the lognormal hypothesis concerning the distribution of energy dissipation in intermittent turbulence,” *Rosenblatt M., Van Atta C. (eds) Statistical Models and Turbulence. Lecture Notes in Physics*, vol. 12, pp. 333–351, 1972.
- [14] U. Frisch and G. Parisi, “Fully developed turbulence and intermittency,” *Turbulence and Predictability in Geophysical Fluid Dynamics and Climate Dynamics*, pp. 84–87, 1983.
- [15] C. MENEVEAU and K. R. SREENIVASAN, “The multifractal spectrum of the dissipation field in turbulent flows,” *Nuclear Physics B*, vol. 2, pp. 49–76, 1987.
- [16] C. Yin, “Encoding and decoding dna sequences by integer chaos game representation,” *Journal of Computational Biology*, vol. 26, no. 2, pp. 143–151, 2019.
- [17] T. Hoang, C. Yin, and S. S.-T. Yau, “Numerical encoding of dna sequences by chaos game representation with application in similarity comparison,” *Genomics*, vol. 108, no. 3-4, pp. 134–142, 2016.
- [18] G. Durán-Meza, J. López-García, and J. L. del Río-Correa, “The Self-similarity Properties and Multifractal Analysis of DNA Sequences,” *Applied Mathematics and Nonlinear Sciences*, vol. 4, no. 1, pp. 267–278, 2019.
- [19] T. Strachan and A. Read, *Human Molecular Genetics*. CRC Press, 2018.
- [20] NIH, “National Human Genome Research Institute.” <https://www.genome.gov/>, 2024.
- [21] P. Deschavanne, A. Giron, J. Vilain, G. Fagot, and B. Fertil, “Genomic Signature: Characterization and Classification of Species Assessed by Chaos Game Representation of Sequences,” *Mol. Biol. Evol.*, vol. 16, pp. 1391–1399, 1999.
- [22] J. S. Almeida, J. A. Carriço, A. Maretzek, P. A. Noble, and M. Fletcher, “Analysis of genomic sequences by chaos game representation,” *Bioinformatics*, vol. 17, no. 5, pp. 429–437, 2001.
- [23] E. Y. Chan and R. M. Corless, “Chaos game representation,” *SIAM Review*, vol. 65, no. 1, pp. 261–290, 2023.
- [24] D. Harte, *Multifractals: Theory and Applications*. Boca Raton, Florida: Chapman and Hall CRC, 2001.

- [25] C. Beck and F. Schögl, *Thermodynamics of Chaotic Systems An Introduction*. Cambridge: Cambridge University Press, 1993.
- [26] J. E. Hutchinson, “Fractals and self similarity,” *Indiana University Mathematics Journal*, vol. 30, no. 05, pp. 713–747, 1981.
- [27] A. W. Juan H, Arredondo R, *Medidas e Integrales*. Ciudad de México, México: Universidad Autónoma Metropolitana, 2010.
- [28] B. O. Koopman, “Hamiltonian systems and transformation in hilbert space,” *Proceedings of the National Academy of Sciences*, vol. 17, no. 5, pp. 315–318, 1931.
- [29] I. Mezić, “Analysis of fluid flows via spectral properties of the koopman operator,” *Annual Review of Fluid Mechanics*, vol. 45, pp. 357–378, 2013.
- [30] S. L. Brunton and J. N. Kutz, *Data-driven science and engineering: Machine learning, dynamical systems, and control*. Cambridge University Press, 2021.
- [31] A. Lasota and M. C. Mackey, *Chaos, fractals, and noise: stochastic aspects of dynamics*, vol. 97. Springer Science & Business Media, 1994.
- [32] M. Budisic, R. Mohr, and I. Mezic, “Applied koopmanism,” *Chaos: An Interdisciplinary Journal of Nonlinear Science*, vol. 22, no. 4, p. 047510, 2012.
- [33] I. Mezić, “Spectral properties of dynamical systems, model reduction and decompositions,” *Non-linear Dynamics*, vol. 41, pp. 309–325, 2005.
- [34] S. E. Otto and C. W. Rowley, “Koopman operators for estimation and control of dynamical systems,” *Annual Review of Control, Robotics, and Autonomous Systems*, vol. 4, no. Volume 4, 2021, pp. 59–87, 2021.
- [35] P. J. Schmid, “Dynamic mode decomposition of numerical and experimental data,” *Journal of fluid mechanics*, vol. 656, pp. 5–28, 2010.
- [36] J. H. Tu, C. W. Rowley, D. M. Luchtenburg, S. L. Brunton, and J. N. Kutz, “On dynamic mode decomposition: theory and applications,” *Journal of Computational Dynamics*, vol. 1, no. 2, pp. 391–421, 2014.
- [37] M. O. Williams, I. G. Kevrekidis, and C. W. Rowley, “A data-driven approximation of the koopman operator: Extending dynamic mode decomposition,” *Journal of Nonlinear Science*, vol. 25, pp. 1307–1346, 2015.

- [38] J. V. José and E. J. Saletan, *Classical Dynamics: A Contemporary Approach*. Cambridge University Press, 1998.
- [39] M. Tabor, *Chaos and Integrability in Nonlinear Dynamics*. Wiley-Interscience, 1st ed., 1989.
- [40] M. V. Berry, “Regularity and chaos in classical mechanics, illustrated by three deformations of a circular billiard,” *European Journal of Physics*, vol. 2, no. 2, pp. 91–102, 1981.
- [41] A. Bäcker, “Numerical aspects of eigenvalue and eigenfunction computations for chaotic quantum systems,” 2002. arXiv:nlin/0204061.
- [42] A. Rényi, “On the foundations of information theory,” *REVIEW OF THE INTERNATIONAL STATISTICAL INSTITUTE*, vol. 33, no. 1, pp. 1–15, 1965.
- [43] C. MENEVEAU and K. R. SREENIVASAN, “Simple multifractal cascade model for fully developed turbulence,” *Physical review letters*, vol. 59, no. 13, pp. 1424–1427, 1987.
- [44] T. Vicsek, *Fractal Growth Phenomena*. World Scientific, 1992.
- [45] A. Khinchin, *Mathematical Foundations of Information Theory*. Dover Publications, 1957.
- [46] F. Martínez, “Generalizing Moran’s Theorem,” 2016. ArXiv:1601.04186v1 (math.DS), 16 Jan 2016.
- [47] M. H. Olyaei, A. Yaghoubi, and M. Yaghoobi, “Predicting protein structural classes based on complex networks and recurrence analysis,” *Journal of theoretical biology*, vol. 404, pp. 375–382, 2016.
- [48] H. F. Löchel, D. Eger, T. Sperlea, and D. Heider, “Deep learning on chaos game representation for proteins,” *Bioinformatics*, vol. 36, no. 1, pp. 272–279, 2020.
- [49] Y. A. Álvarez Ballesteros, M. A. Quiroz-Juarez, J. L. Del-Río-Correa, and A. M. Escobar-Ruiz, “Exploring the multifractal behavior of the human genome t2t-chm13v2.0: Graphical representations and cytogenetics,” *Computers in Biology and Medicine*, vol. 197, p. 110977, 2025.
- [50] Y. Wang, K. Hill, S. Singh, and L. Kari, “The spectrum of genomic signatures: from dinucleotides to chaos game representation,” *Gene*, vol. 346, pp. 173–185, 2005.
- [51] H. F. Löchel and D. Heider, “Chaos game representation and its applications in bioinformatics,” *Computational and structural biotechnology journal*, vol. 19, pp. 6263–6271, 2021.
- [52] R. Ren, H. Yu, J. Teng, S. Mao, Z. Bian, Y. Tao, and S. S.-T. Yau, “Cape: a deep learning framework with chaos-attention net for promoter evolution,” *Briefings in Bioinformatics*, vol. 25, no. 5, p. bbae398, 2024.

- [53] A. Rényi, “On the dimension and entropy of probability distributions,” *Acta Mathematica*, vol. 10, pp. 193–215, 1959.



## Apéndice A

# Representación Genómica Binaria de los Cromosomas para Diferentes Valores de $M$

En el cuerpo principal de este trabajo se presentaron los resultados de la Representación Genómica Binaria *BGR* para los valores  $M = 10$  y  $M = 12$ . Si bien estas representaciones permiten caracterizar adecuadamente la distribución de las subsecuencias y revelar propiedades estructurales relevantes del genoma, resulta de interés examinar cómo dichas representaciones evolucionan al considerar diferentes valores de  $M$ .

Por esta razón, en el presente apéndice se muestran las representaciones *BGR* y los espectros multifractales correspondientes para distintos valores de  $M$ , analizados cromosoma por cromosoma. Este estudio complementario permite evaluar la influencia de la longitud de las subsecuencias sobre la estructura geométrica y multifractal observada, así como comparar de manera más detallada las características de los distintos cromosomas del genoma humano.

### A.1. Comparación entre Cromosomas

En las figuras siguientes se presentan las representaciones genómicas binarias *BGR* correspondientes a los cromosomas humanos para valores de  $M$  comprendidos entre 1 y 12. Estas representaciones permiten realizar una comparación directa de los patrones de distribución asociados a cada cromosoma a diferentes escalas de resolución.

El análisis conjunto de las *BGR* facilita la identificación de similitudes y diferencias en la organización de las subsecuencias genómicas, así como la evaluación de la influencia del parámetro  $M$  sobre la estructura observada. De esta manera, es posible examinar cómo evolucionan las características geométricas de las distribuciones y determinar si existen rasgos distintivos o comportamientos comunes entre los distintos cromosomas del genoma humano.

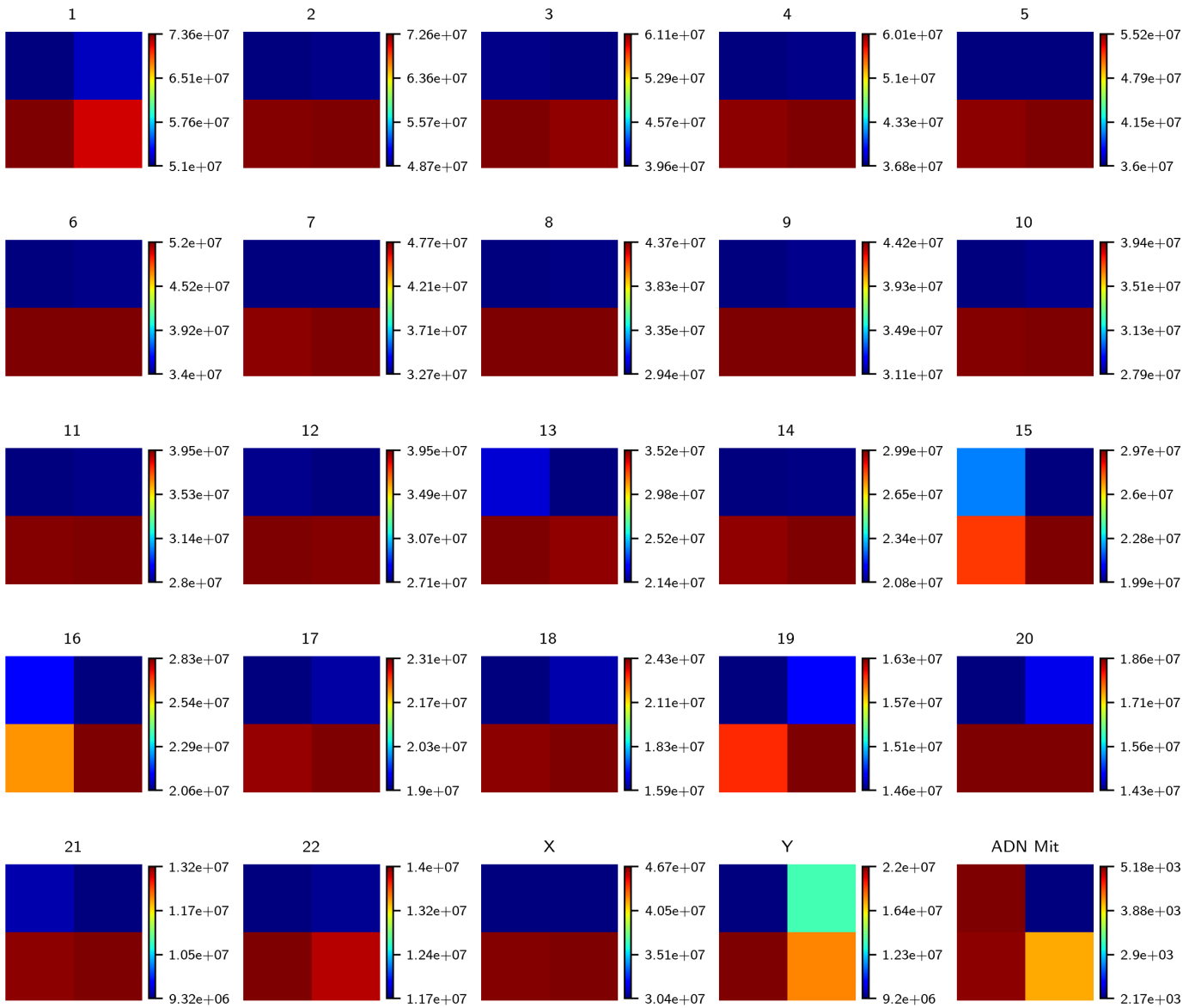


Figura A.1:  $M = 1$



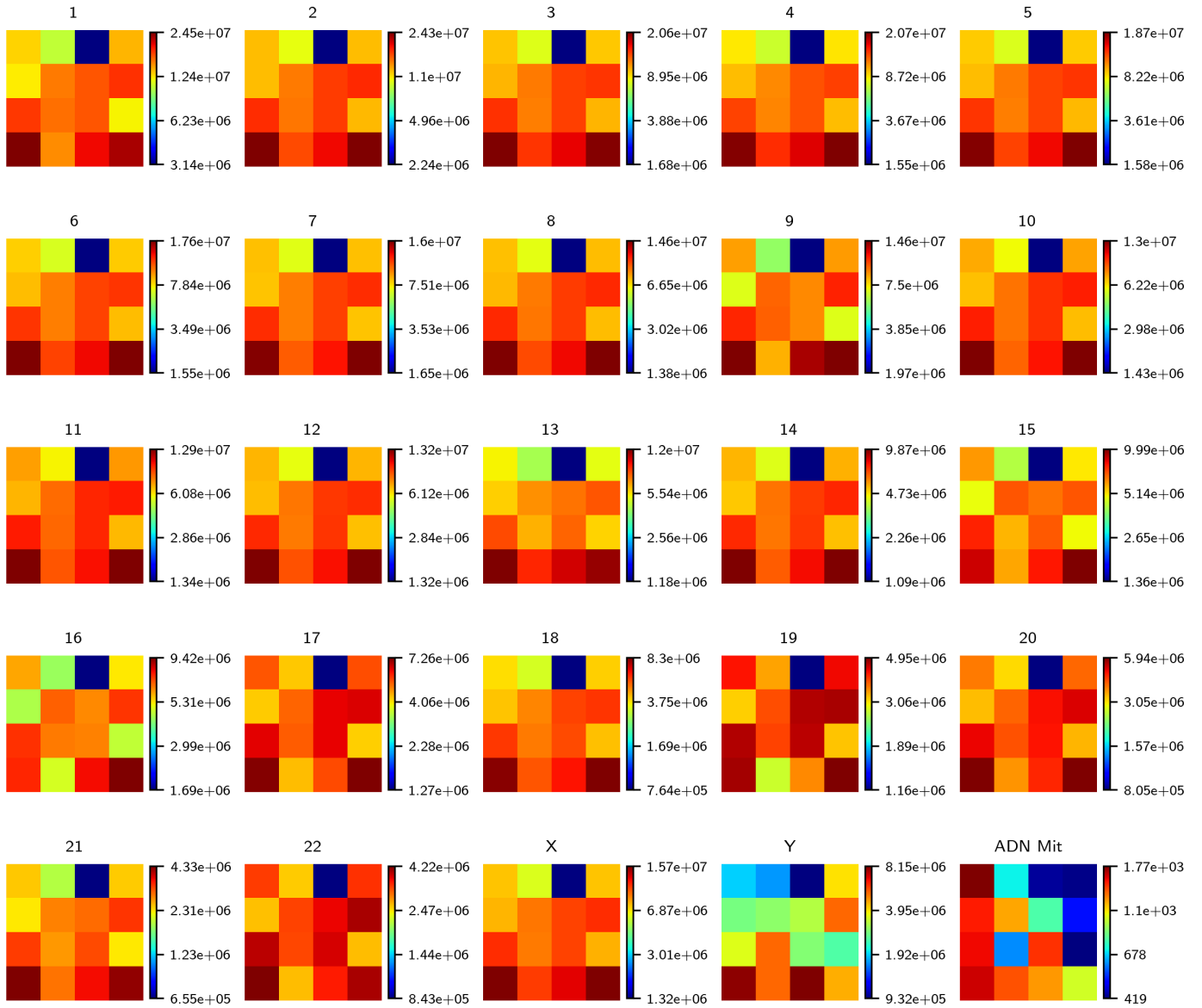


Figura A.2:  $M = 2$

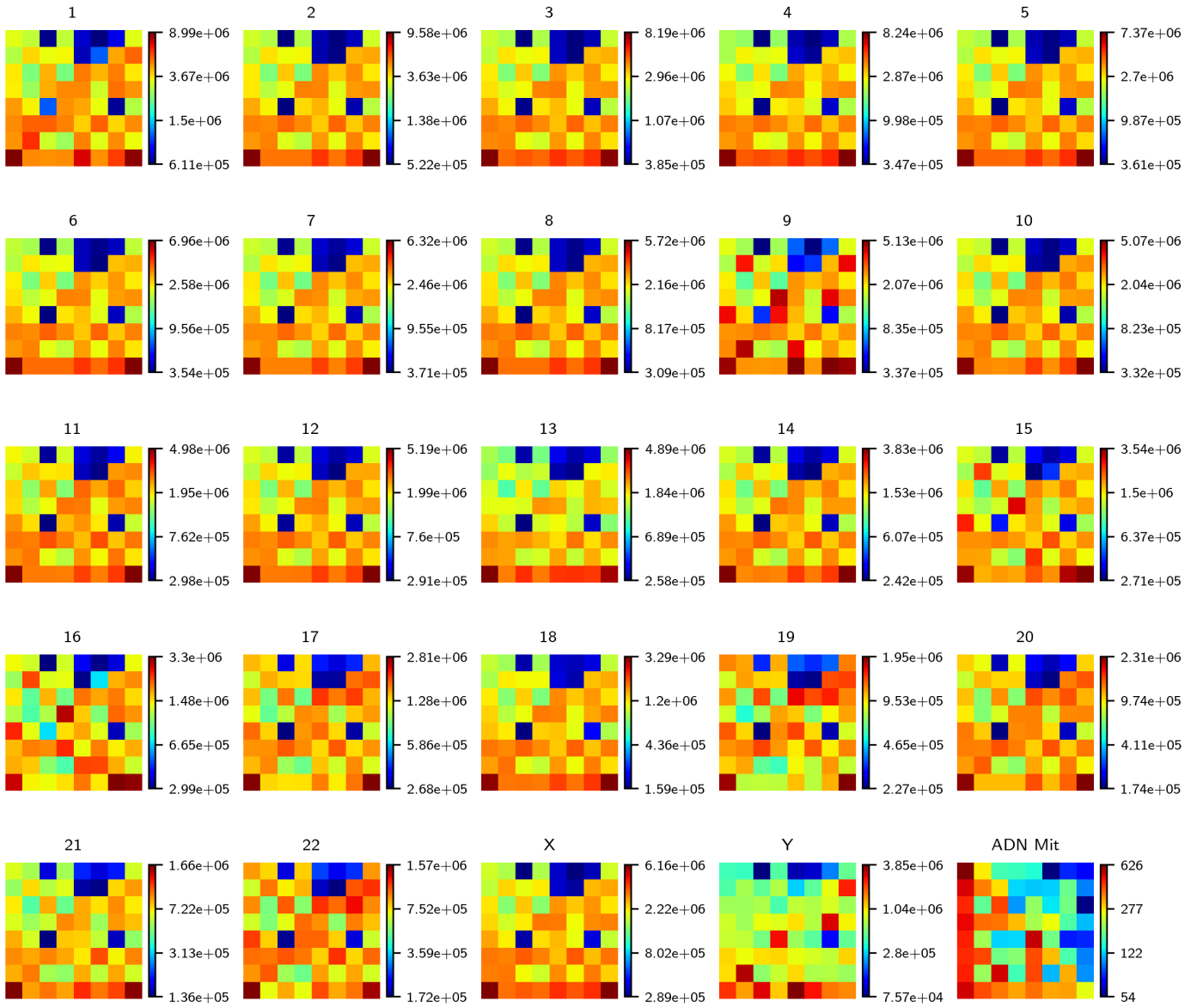


Figura A.3:  $M = 3$

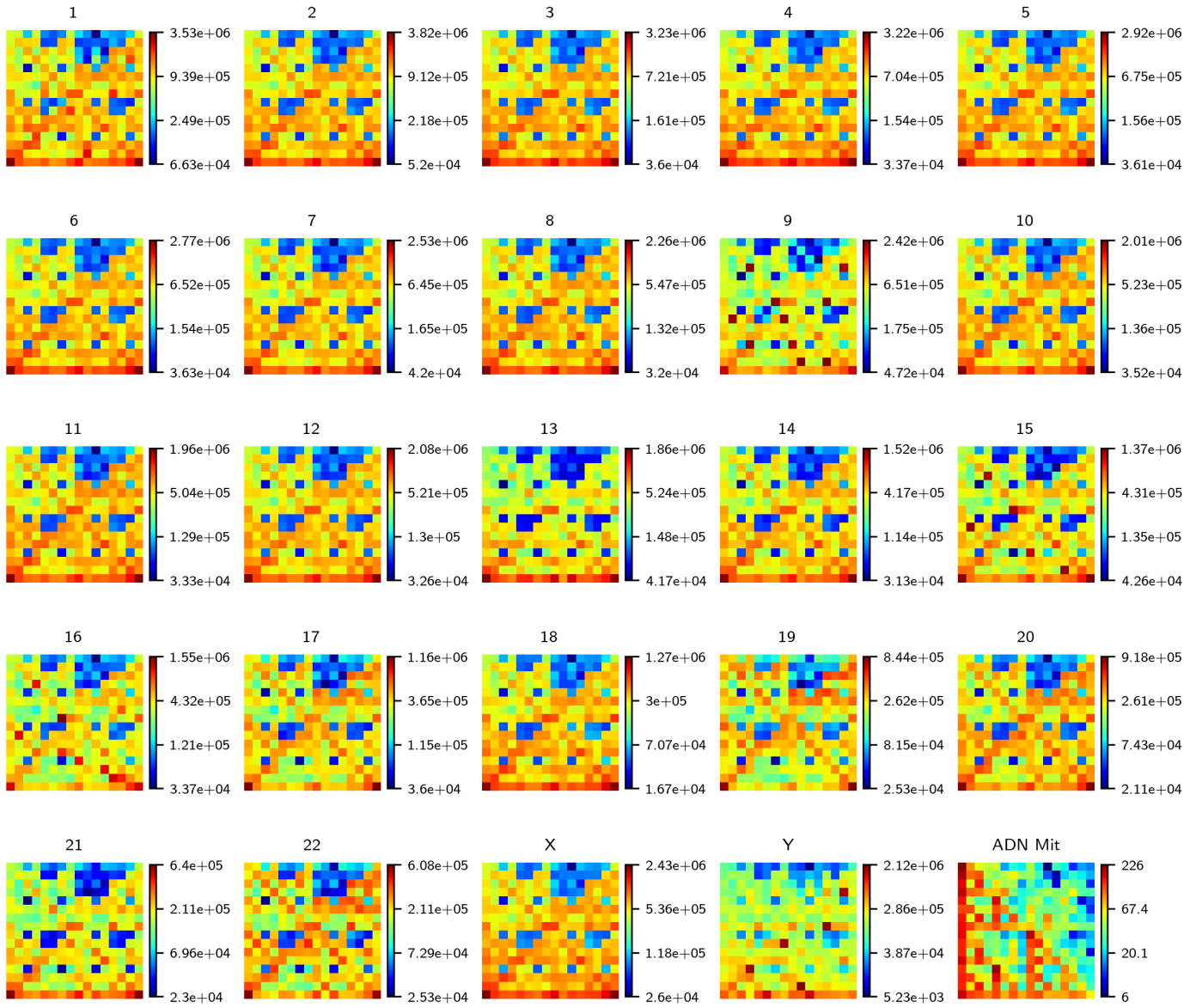


Figura A.4:  $M = 4$

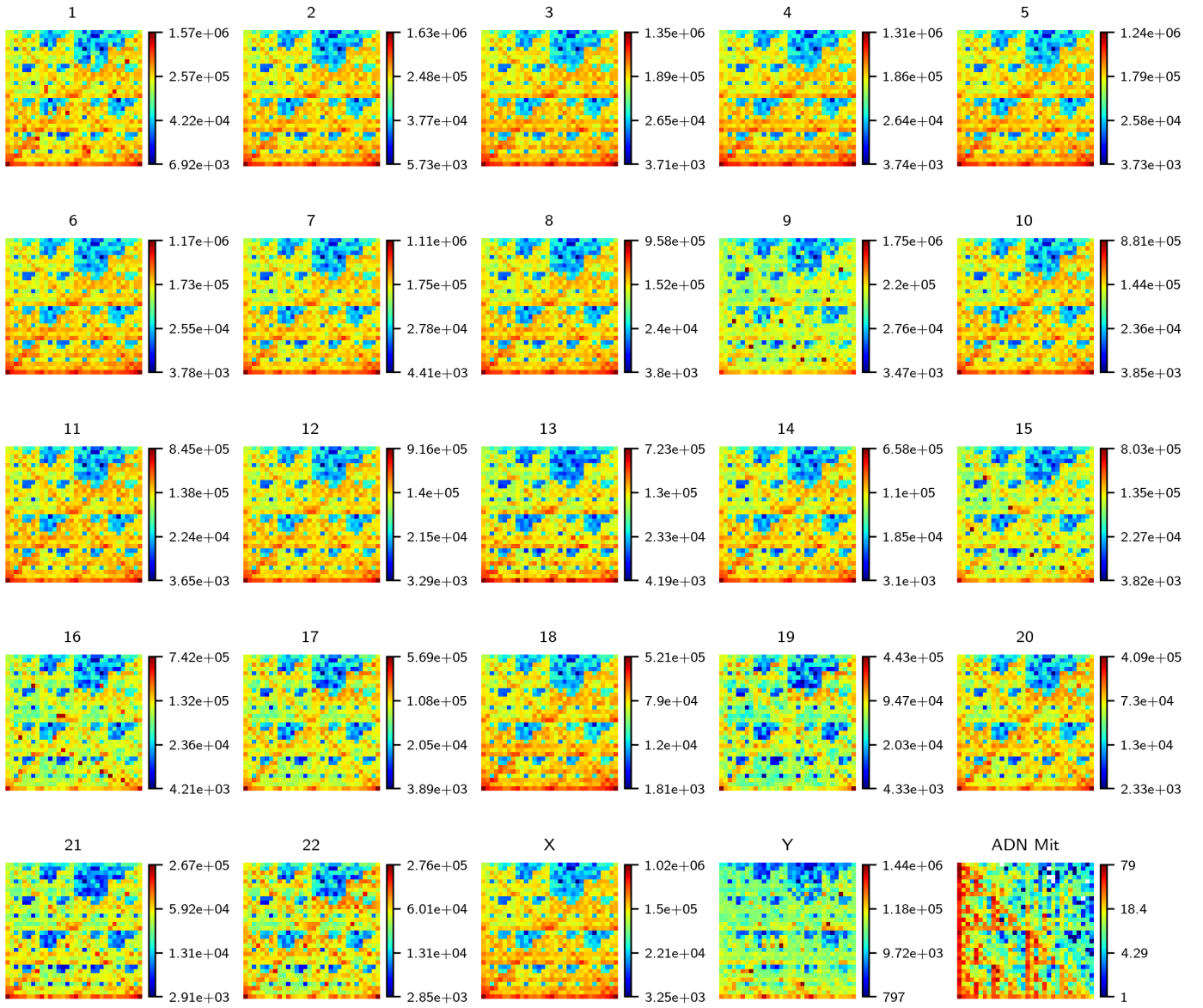


Figura A.5:  $M = 5$

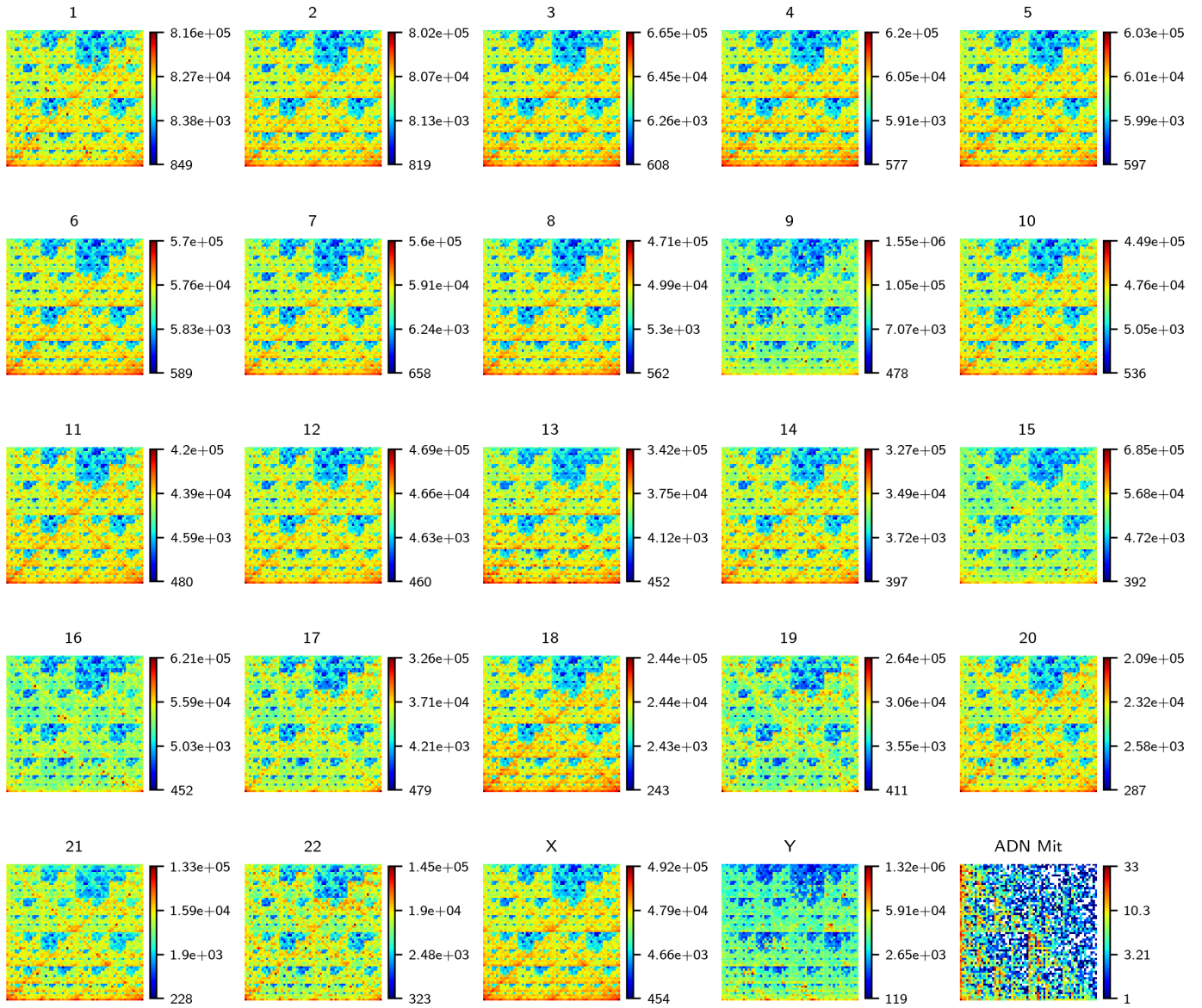


Figura A.6:  $M = 6$



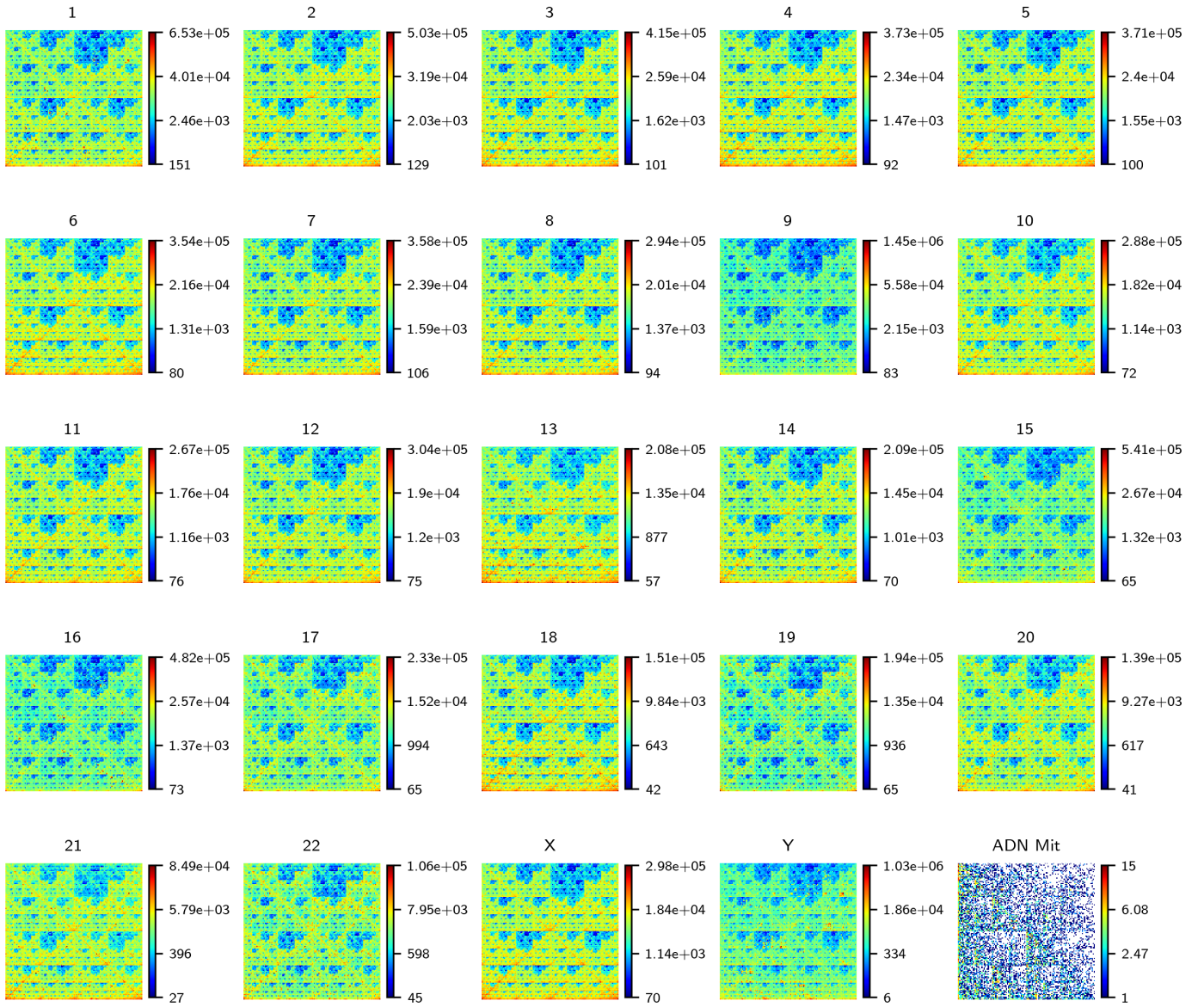


Figura A.7:  $M = 7$

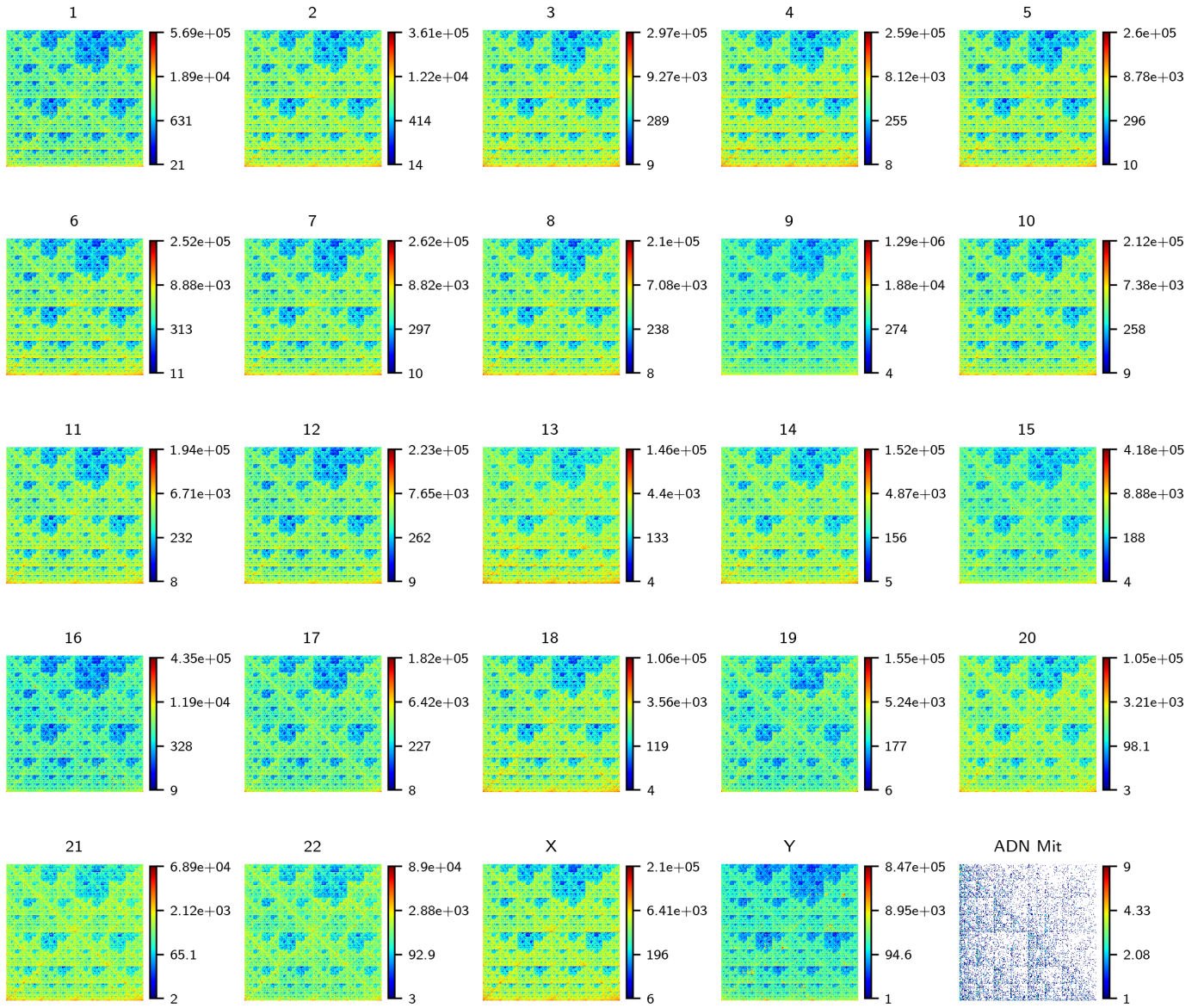


Figura A.8:  $M = 8$



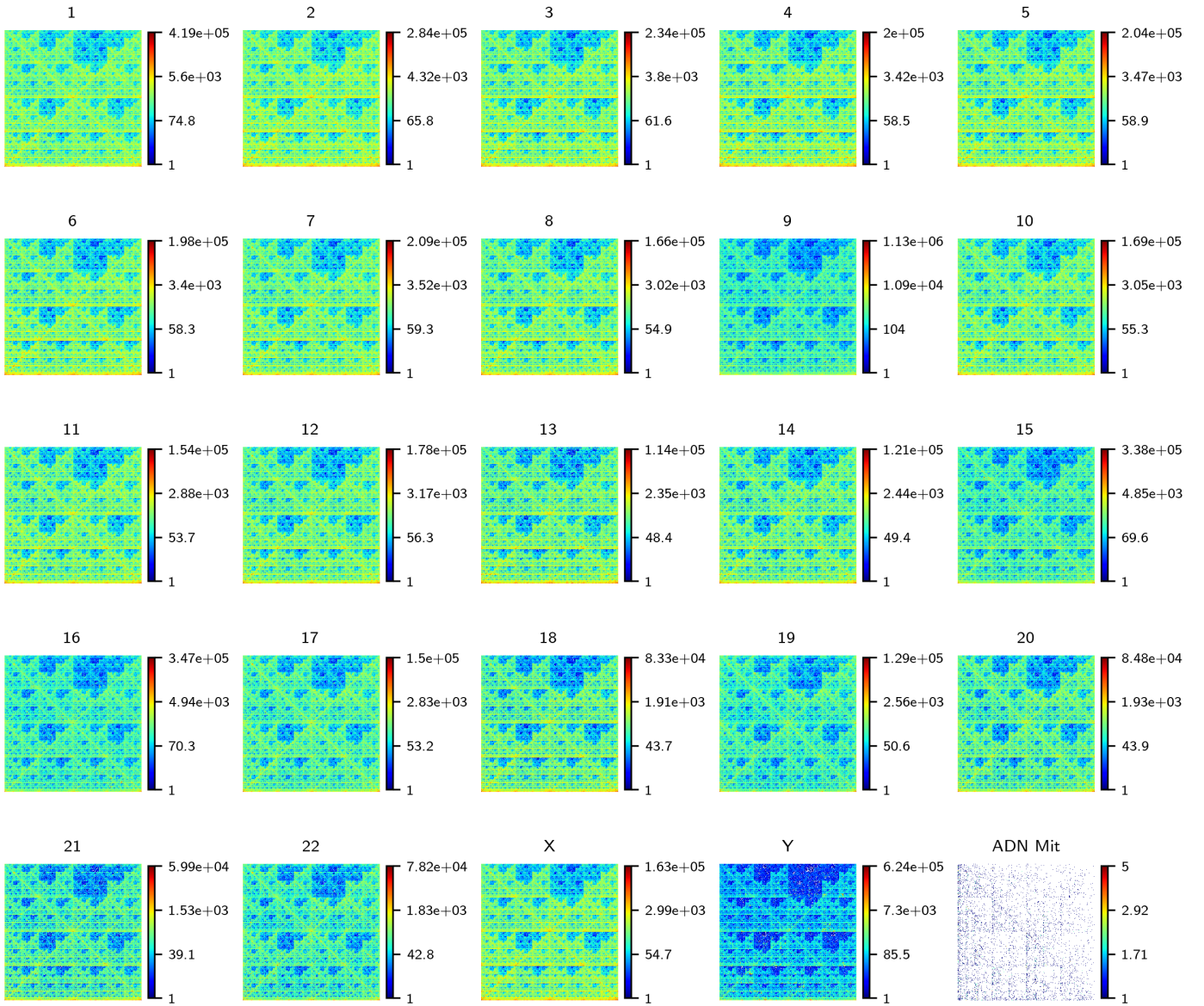


Figura A.9:  $M = 9$



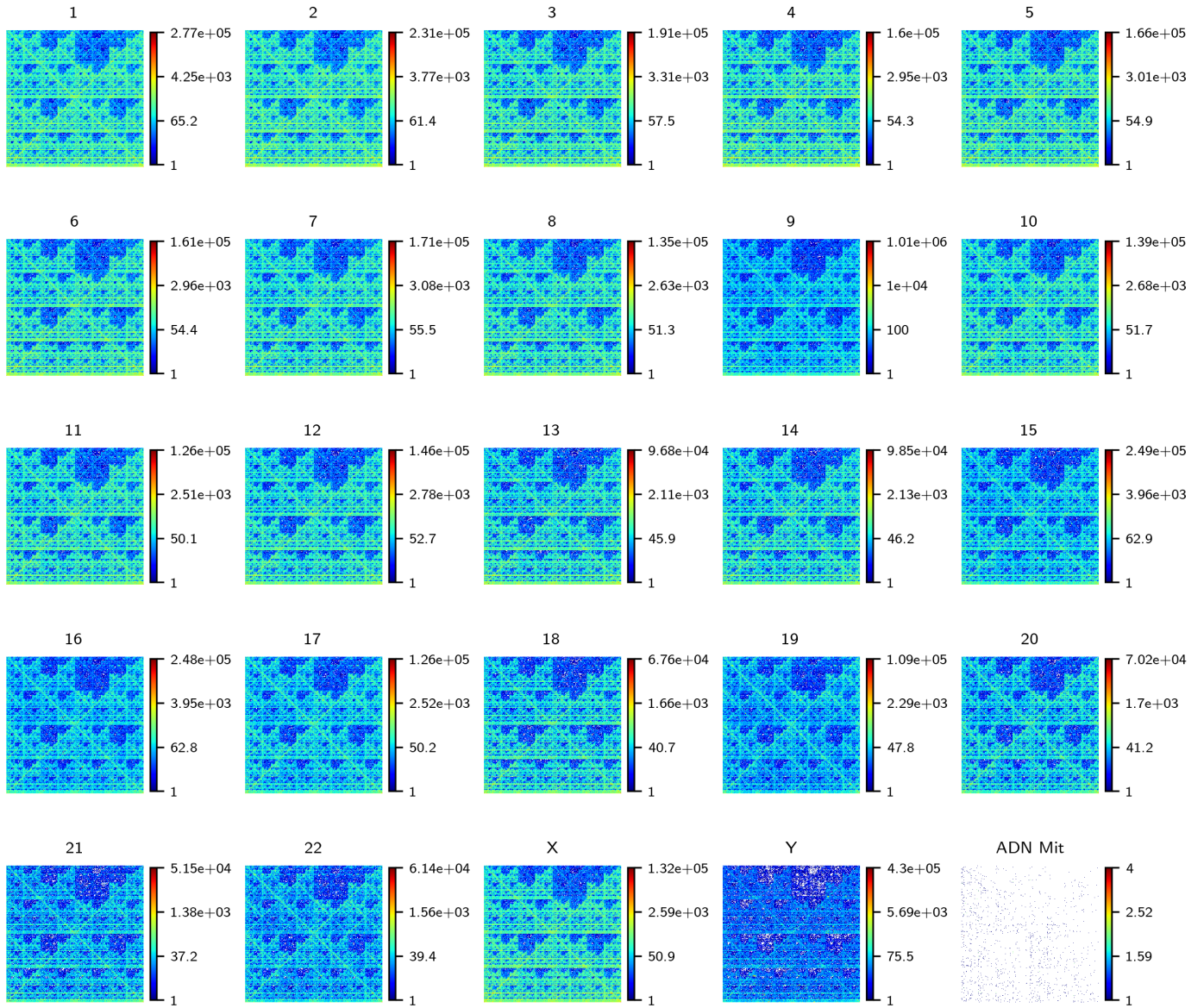


Figura A.10:  $M = 10$

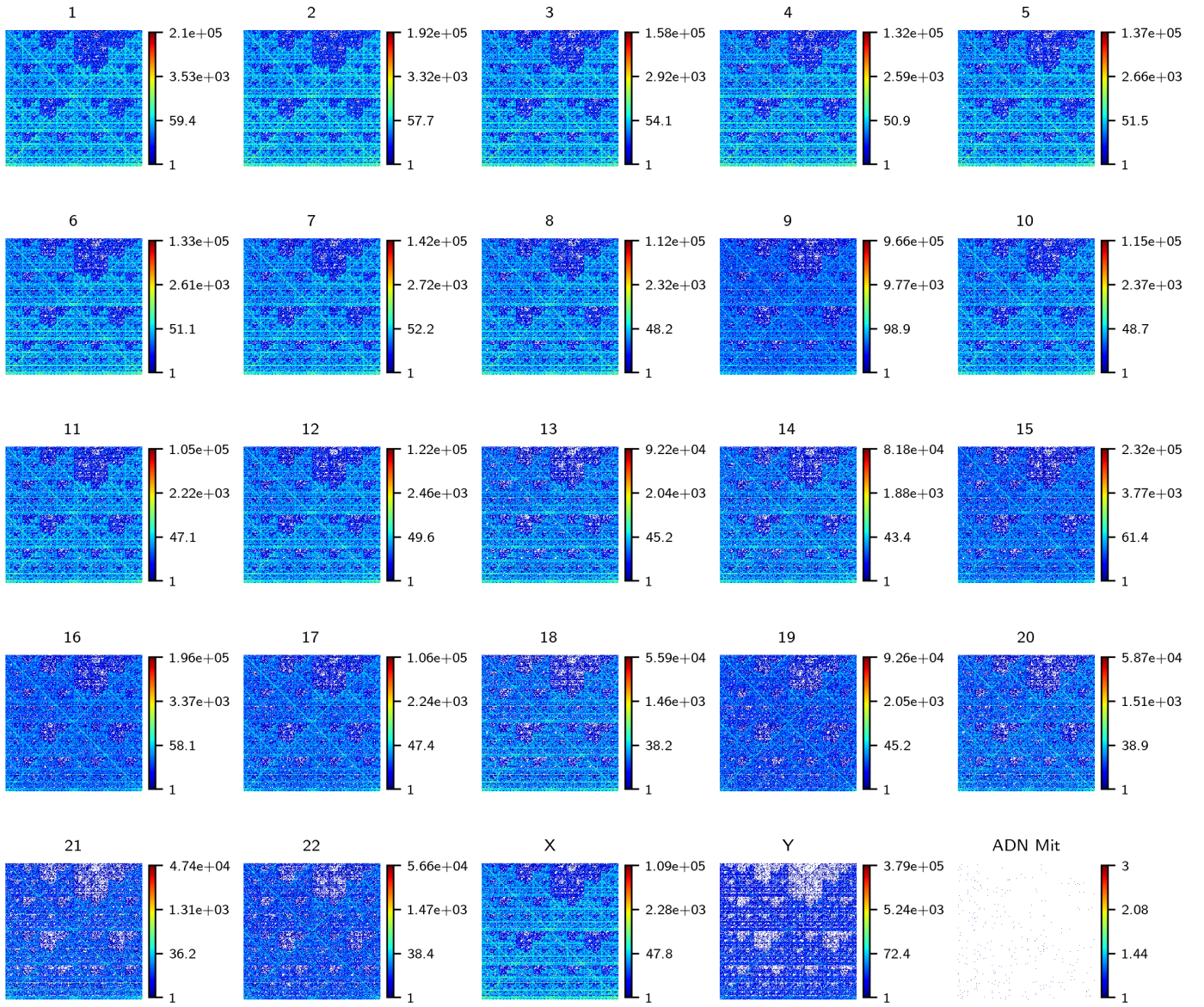


Figura A.11:  $M = 11$



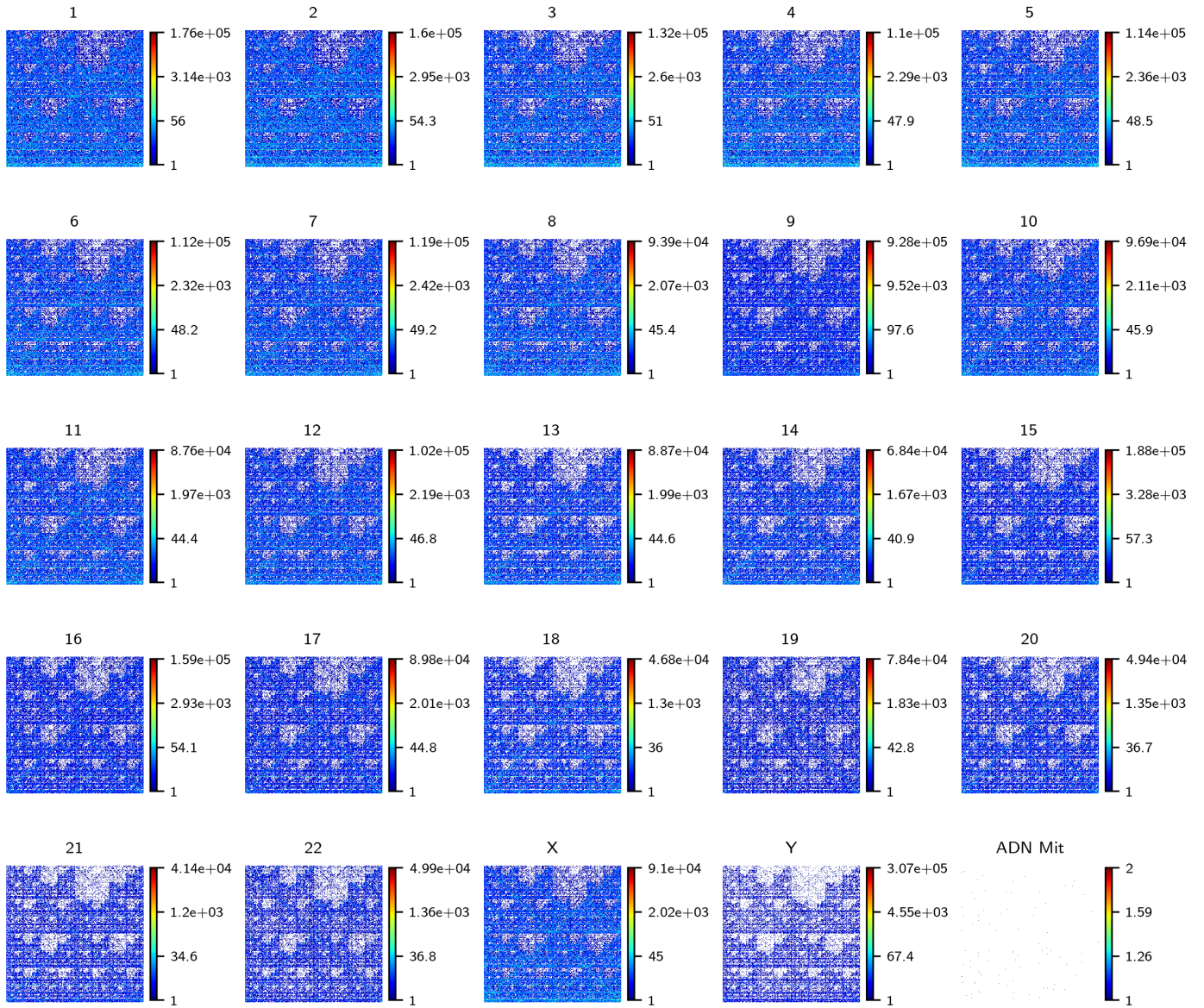


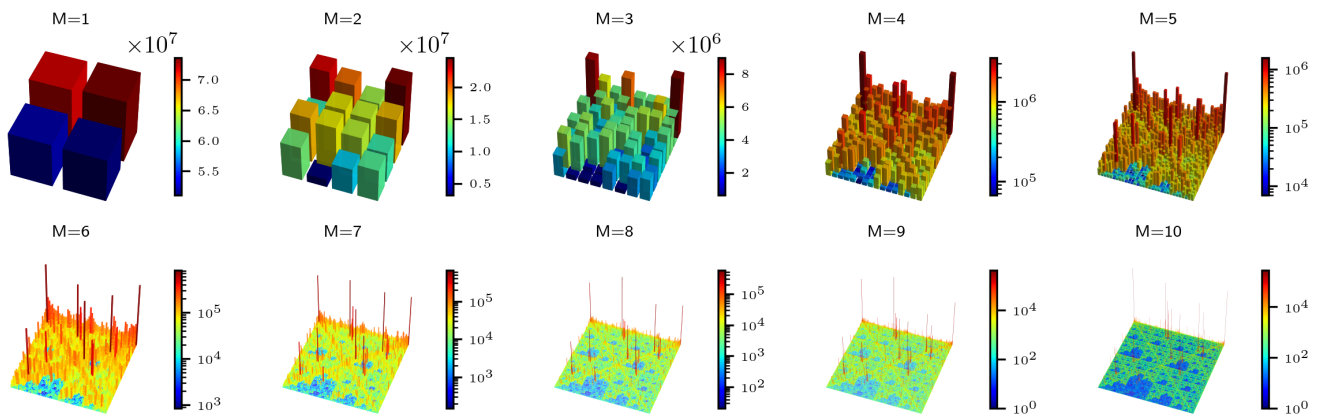
Figura A.12:  $M = 12$

## A.2. Representaciones Tridimensionales por Cromosoma

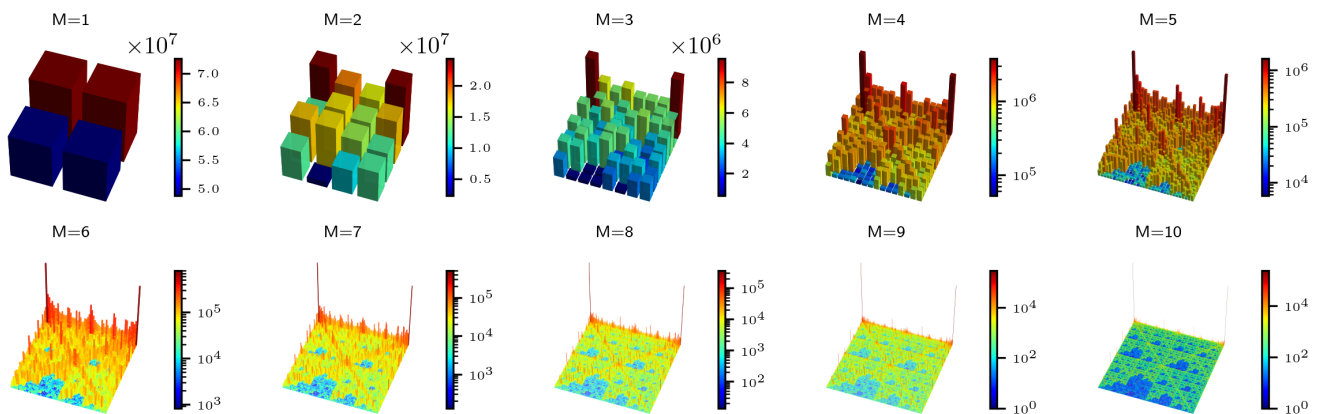
En las figuras siguientes se presentan las representaciones tridimensionales de cada cromosoma para valores de  $M$  comprendidos entre 1 y 10. Estas visualizaciones permiten examinar cómo evoluciona la distribución espacial de las subsecuencias a medida que aumenta su longitud, proporcionando una perspectiva complementaria a las representaciones bidimensionales discutidas en el cuerpo principal del trabajo.

Asimismo, la comparación entre los distintos valores de  $M$  facilita la identificación de patrones estructurales y características geométricas que emergen a diferentes escalas, permitiendo una apreciación más detallada de la organización de cada cromosoma dentro del marco de la Representación Genómica Binaria.

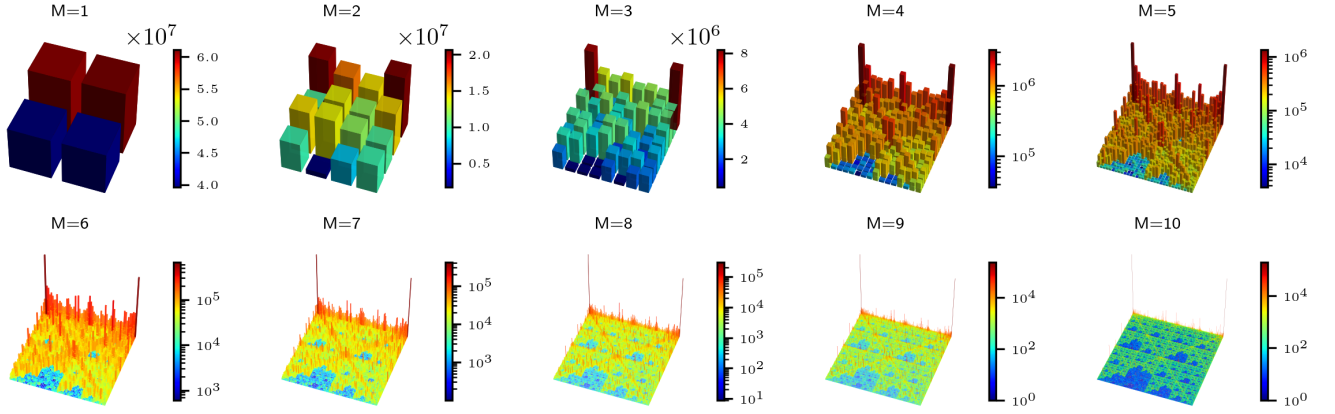
1



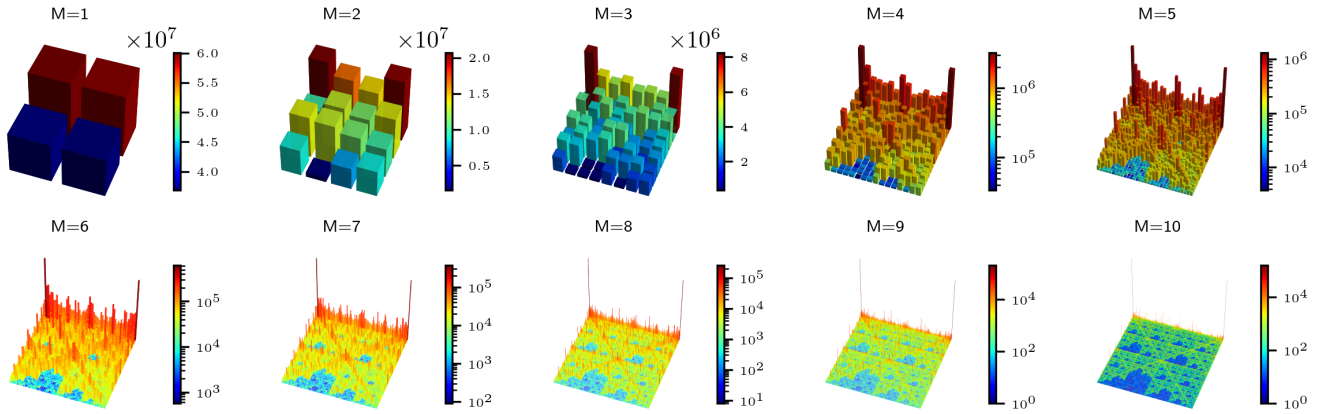
2



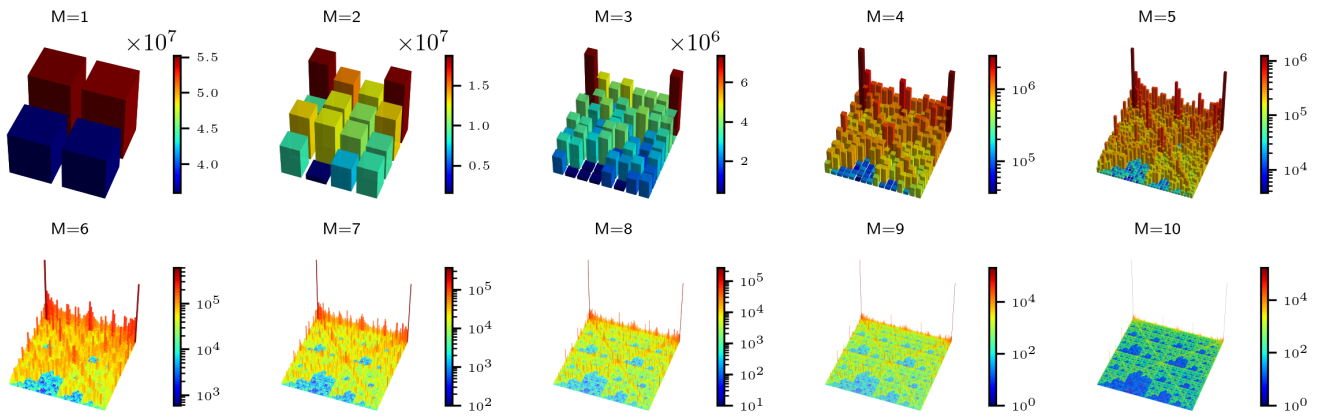
3



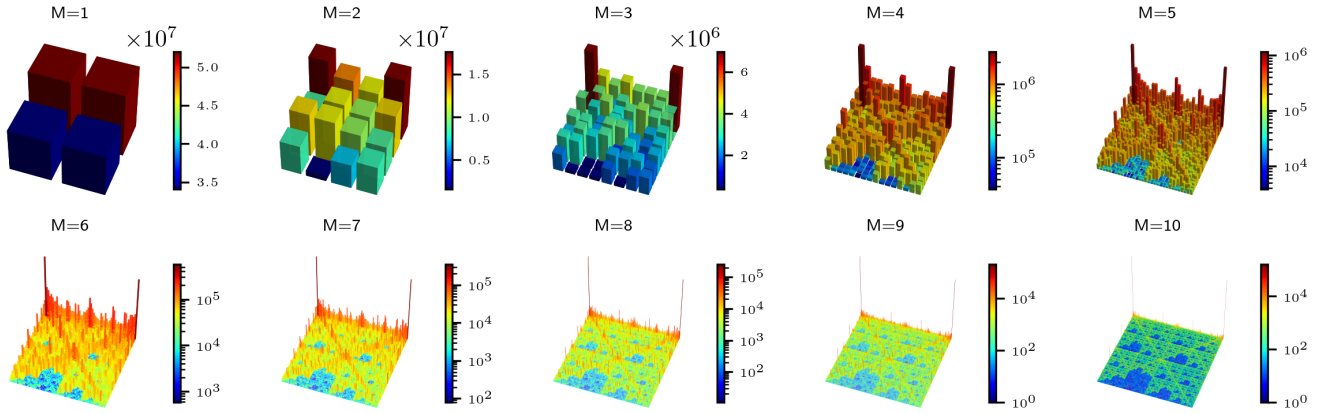
4



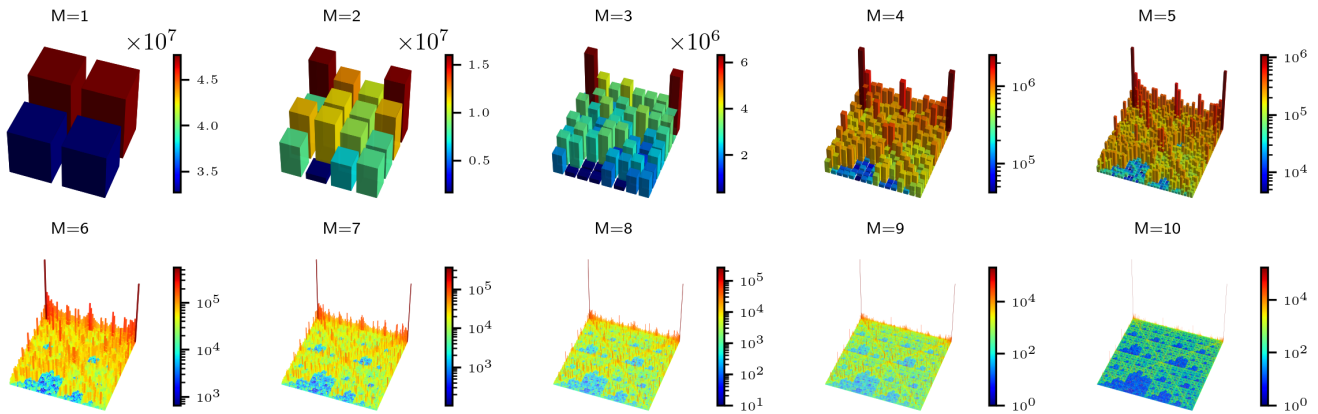
5



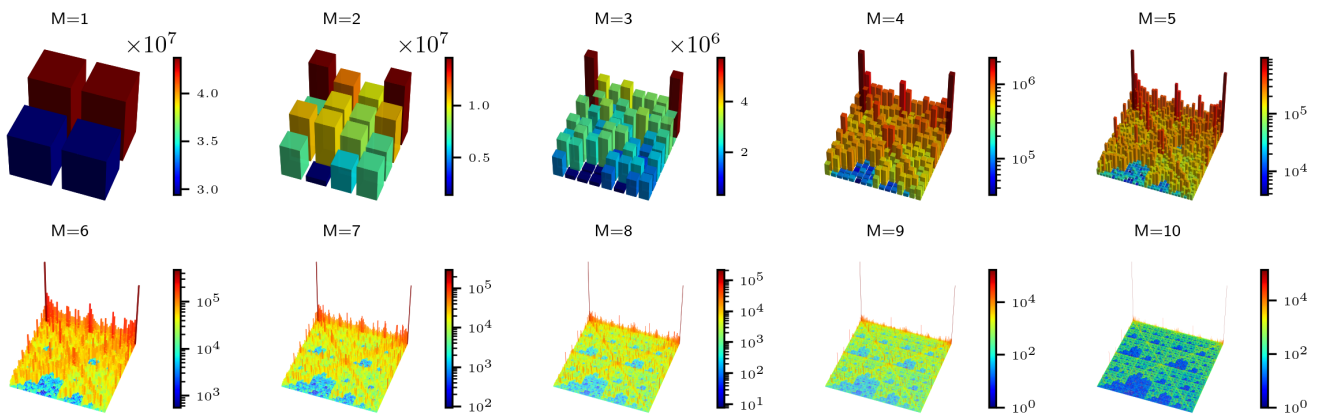
6



7

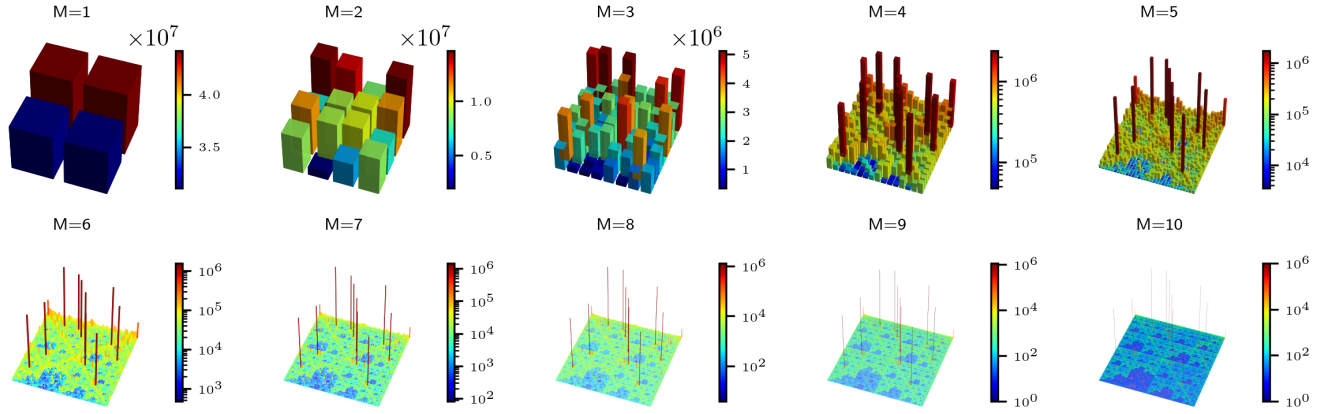


8

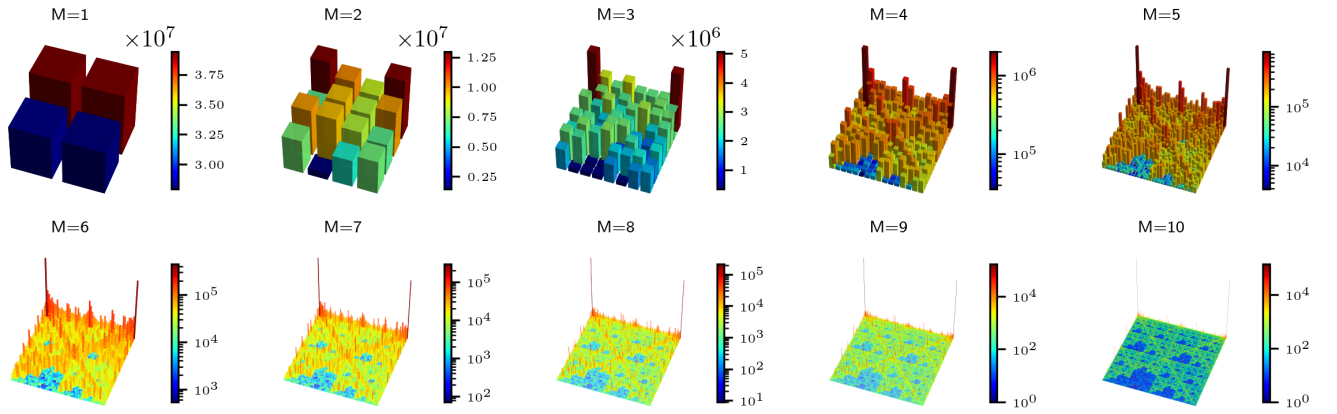




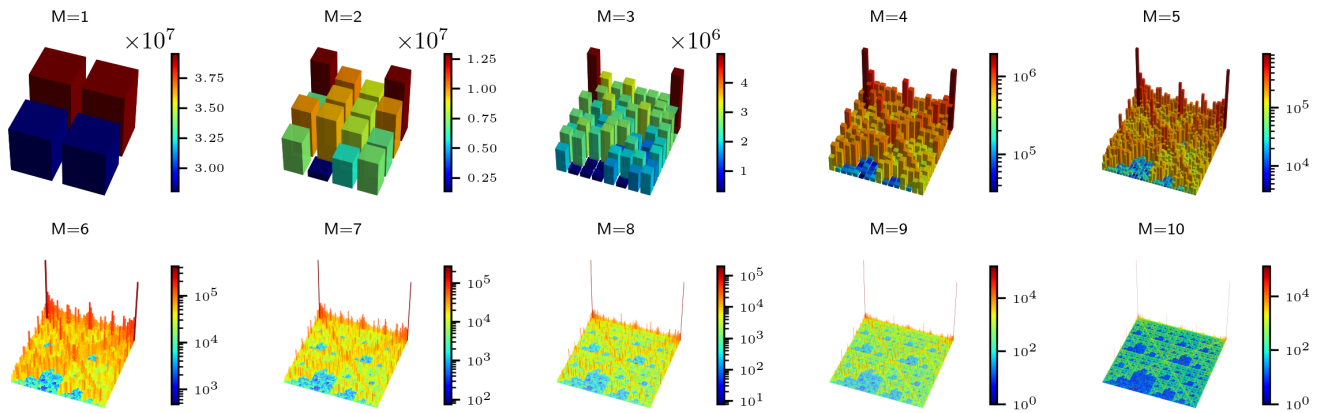
9



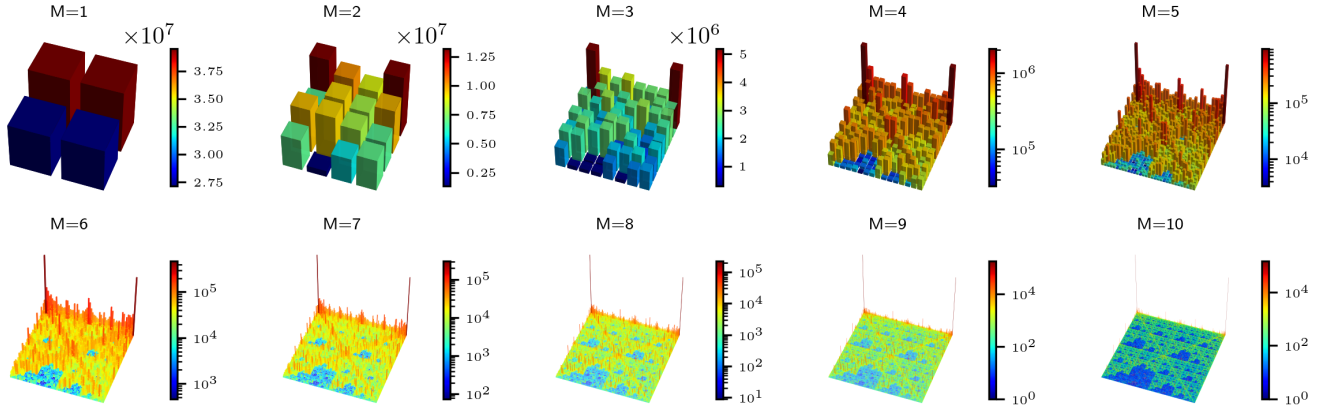
10



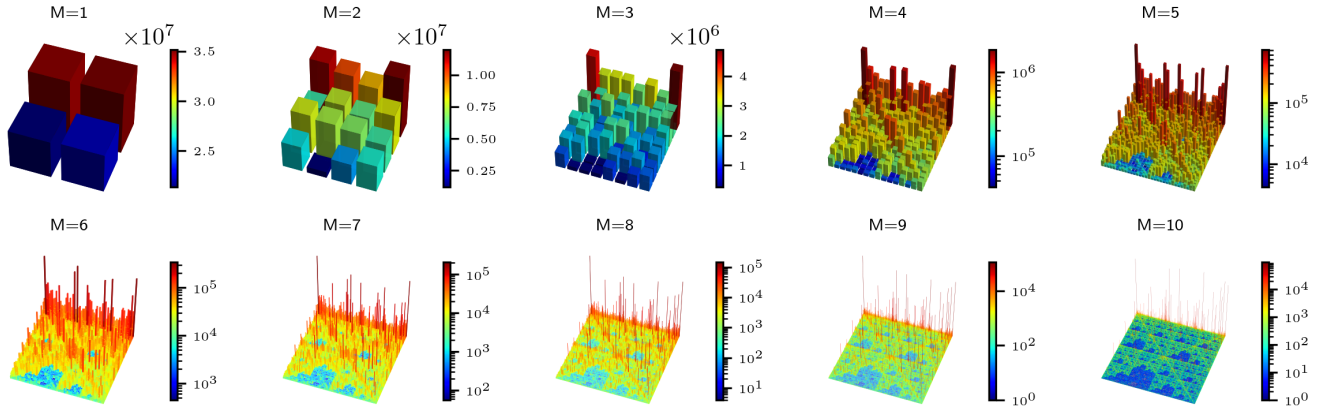
11



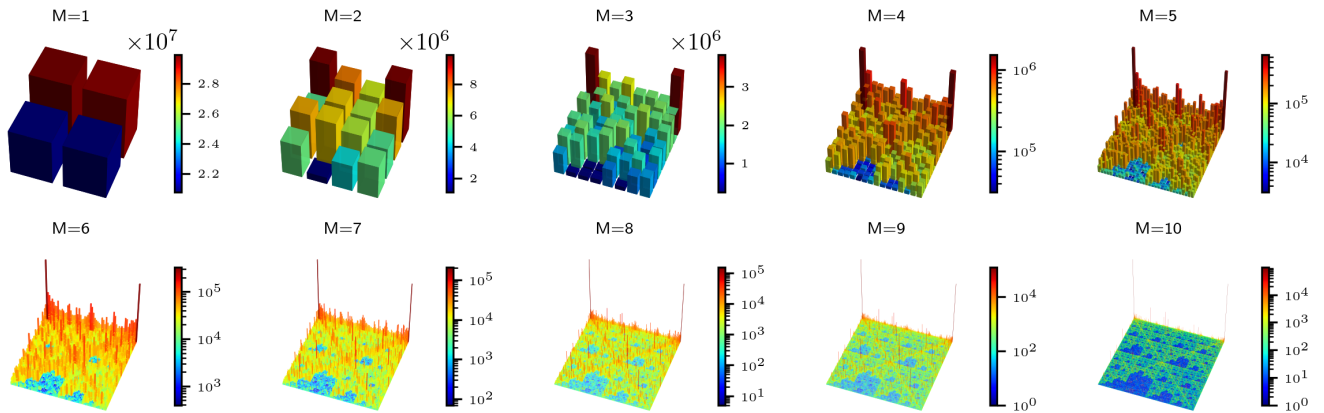
12



13

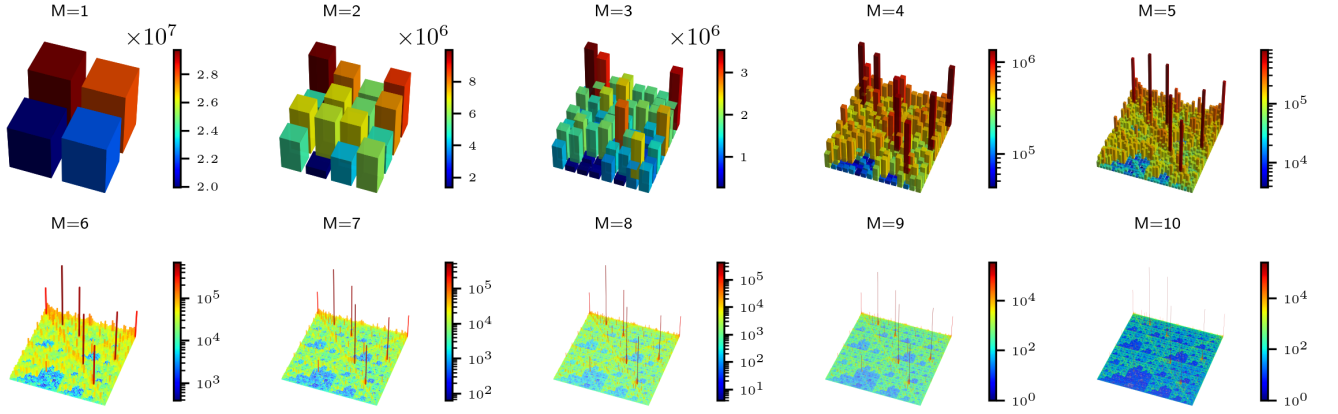


14

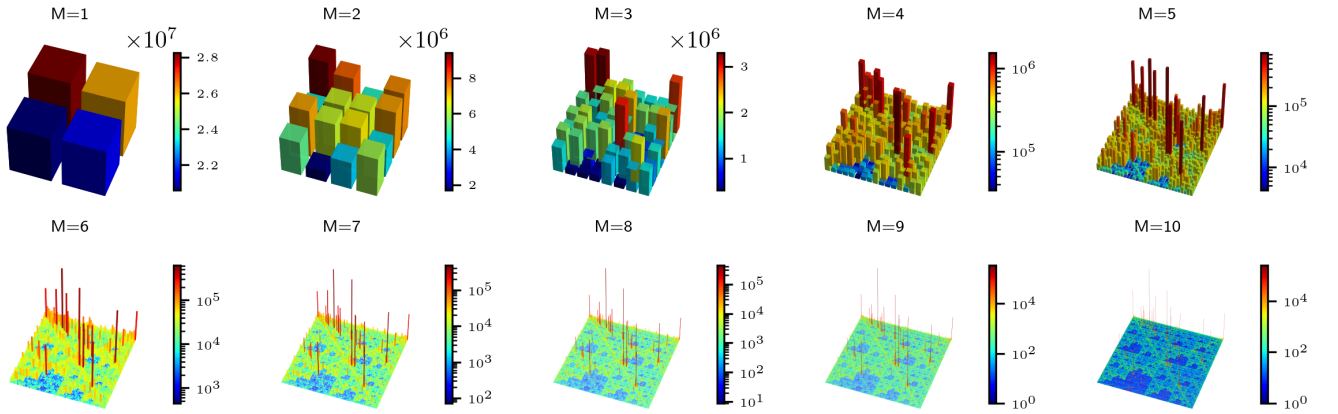




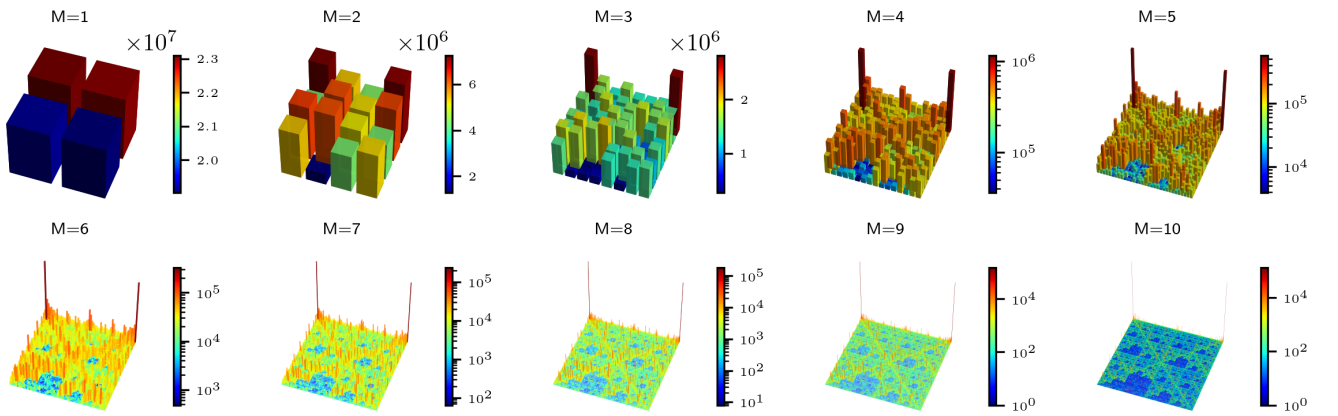
15



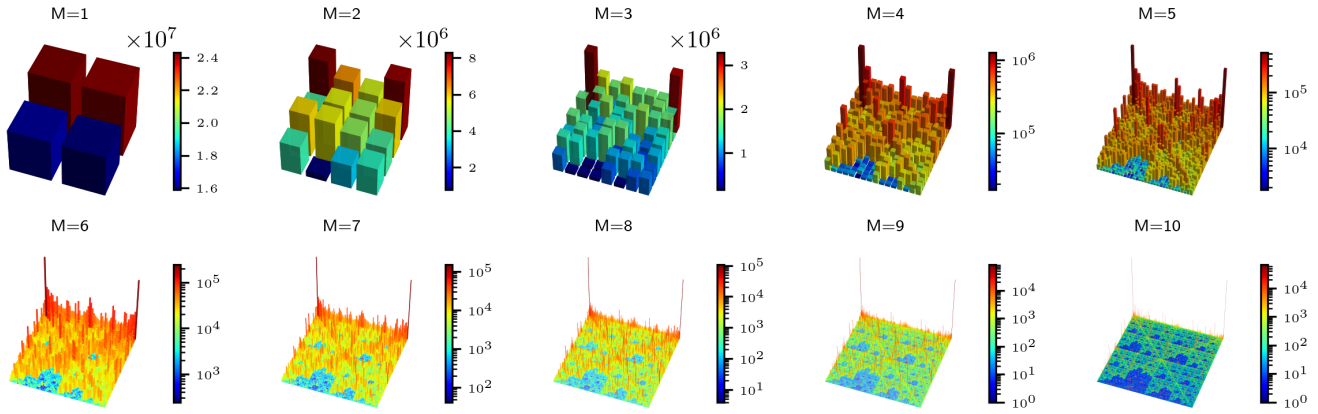
16



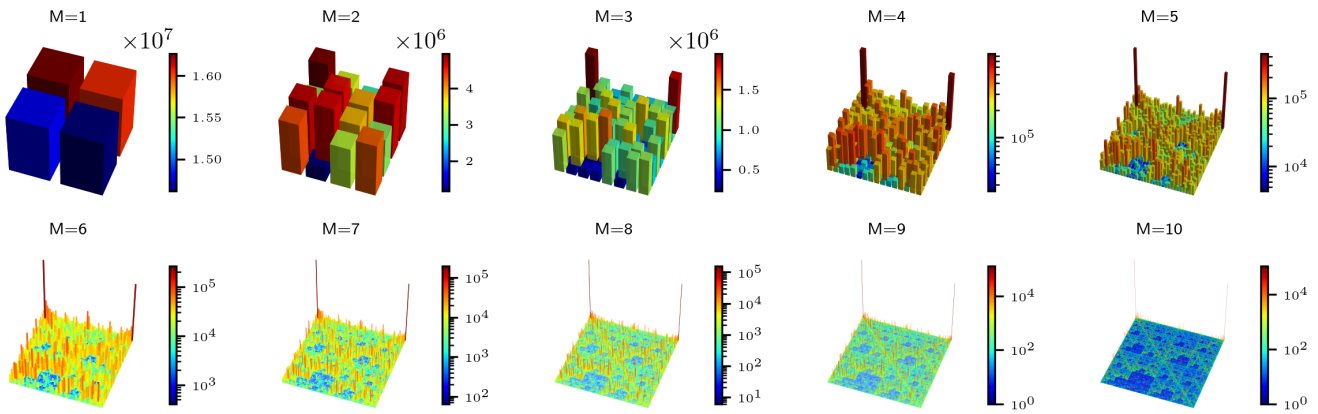
17



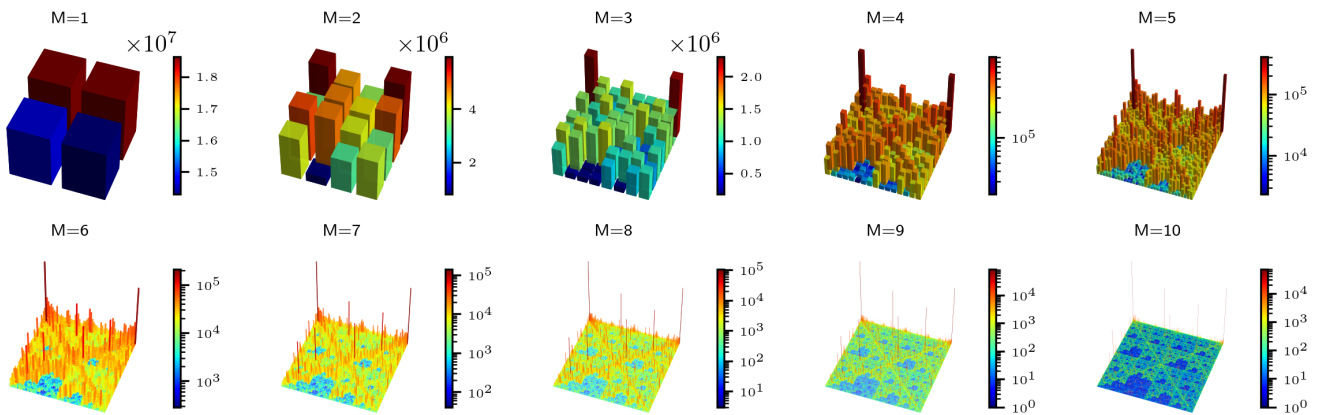
18



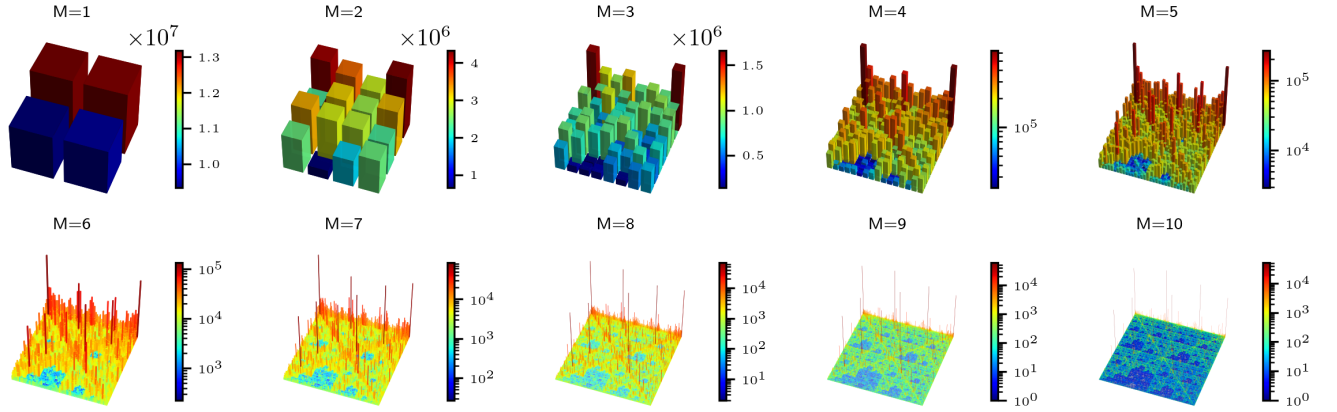
19



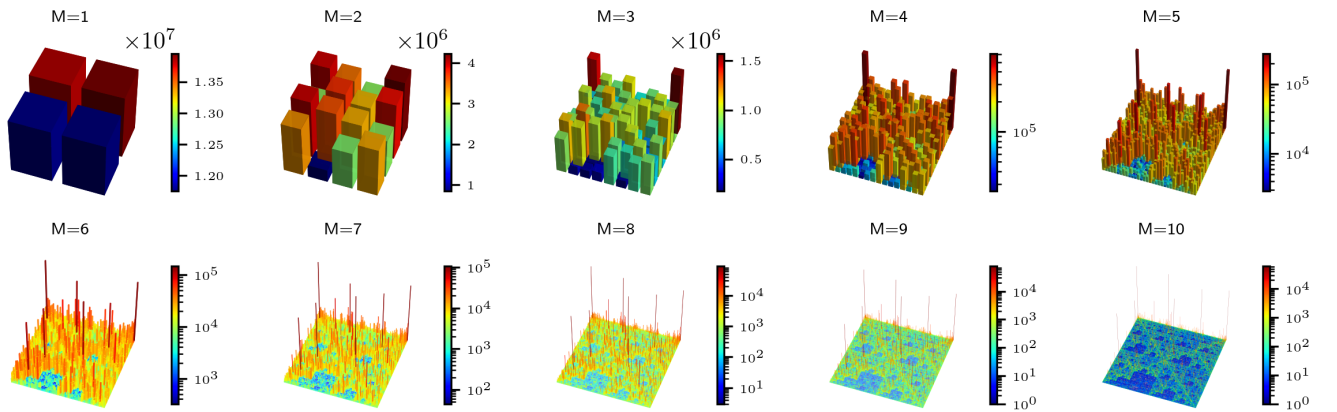
20



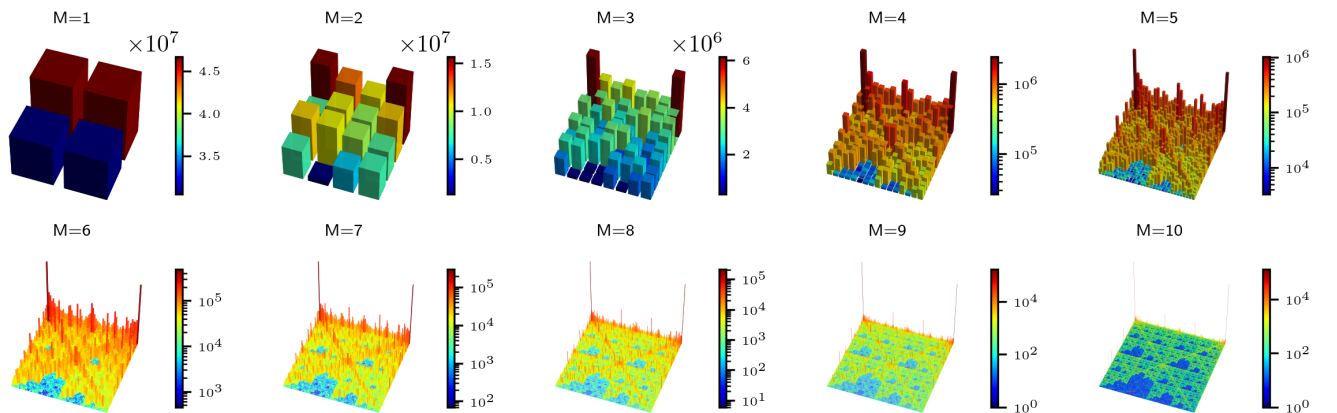
21

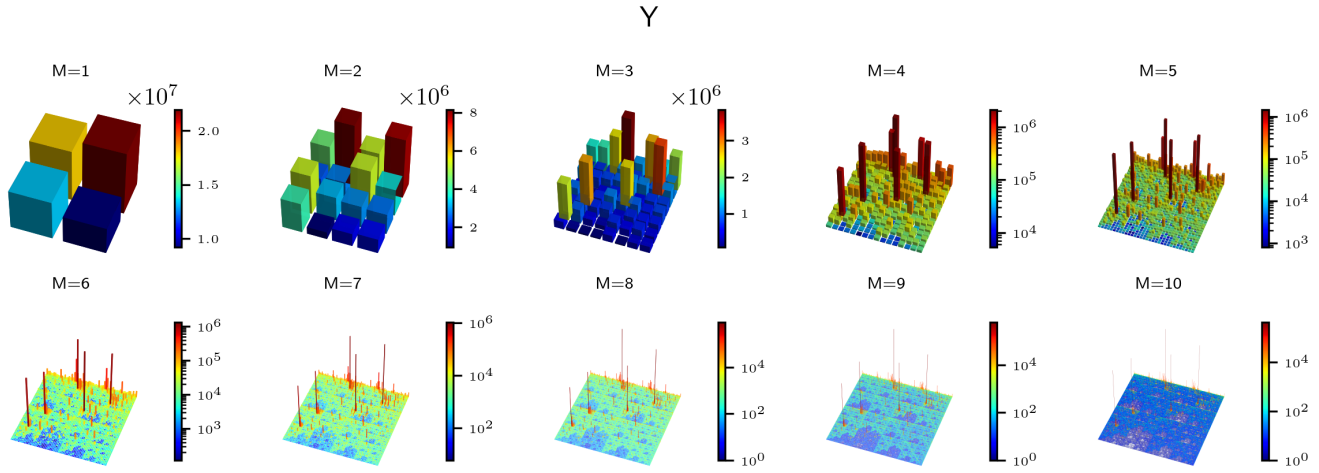


22



X





### A.3. Espectros Multifractales

En las dos figuras siguientes se muestran los espectros multifractales de cada cromosoma para distintos valores de  $M$ . En primer lugar, los resultados se presentan de manera individual para cada cromosoma y, posteriormente, se realiza una comparación entre los espectros de todos ellos. Estas representaciones permiten observar cómo los espectros varían con el valor de  $M$ , parámetro que puede interpretarse como una modificación en el tamaño de caja o cubierta utilizada en la representación.

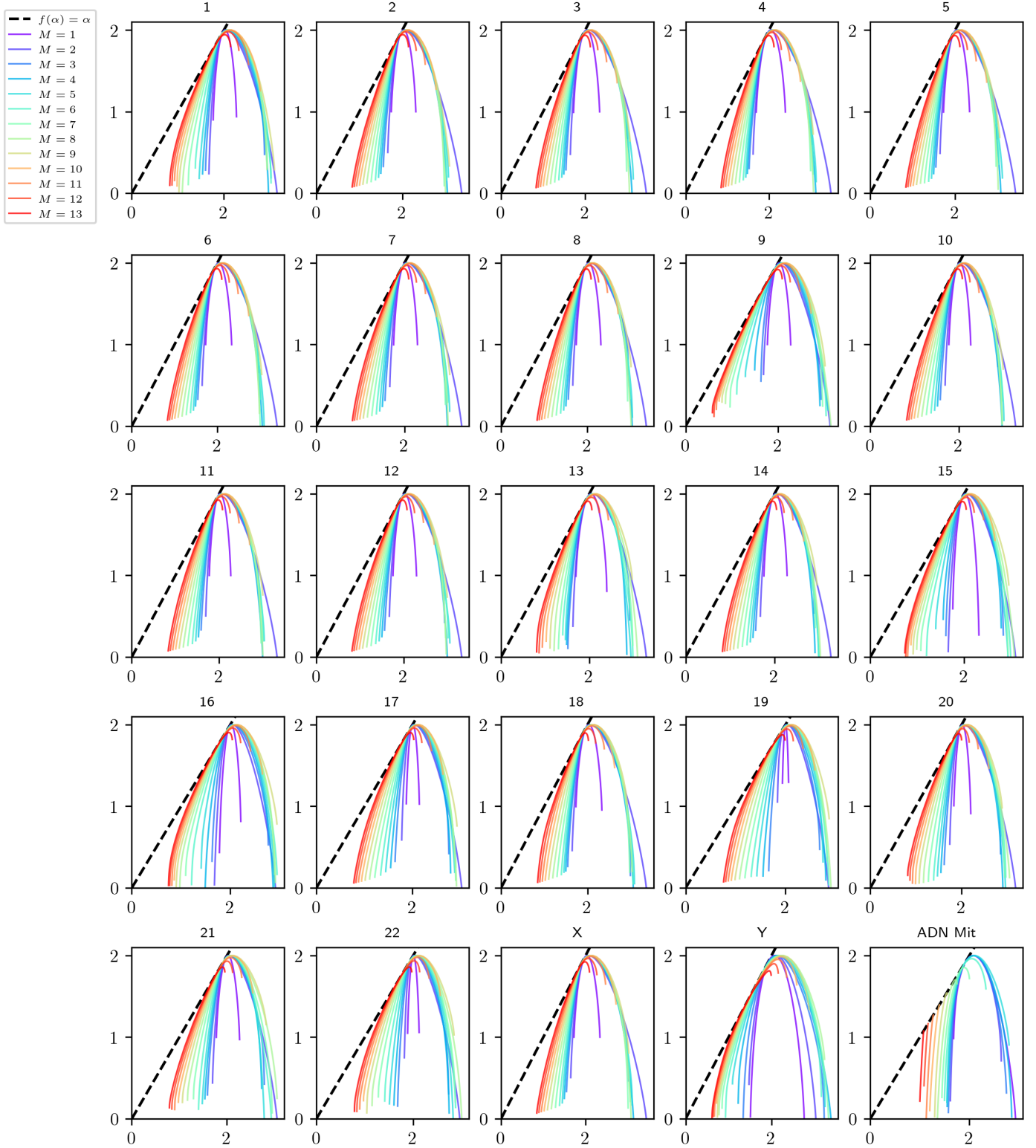
Para los cálculos realizados, se emplearon contracciones sucesivas de razón  $1/2$ , partiendo de una partición inicial del intervalo. Esta configuración garantiza una cubierta óptima bajo la hipótesis de que la distancia mínima entre dos puntos se alcanza cuando el nucleótido correspondiente se repite en el paso siguiente de la construcción, evitando así la intersección entre cubiertas. Como consecuencia, el tamaño de la cubierta depende directamente del valor de  $M$ , es decir, de la longitud de las subsecuencias consideradas.

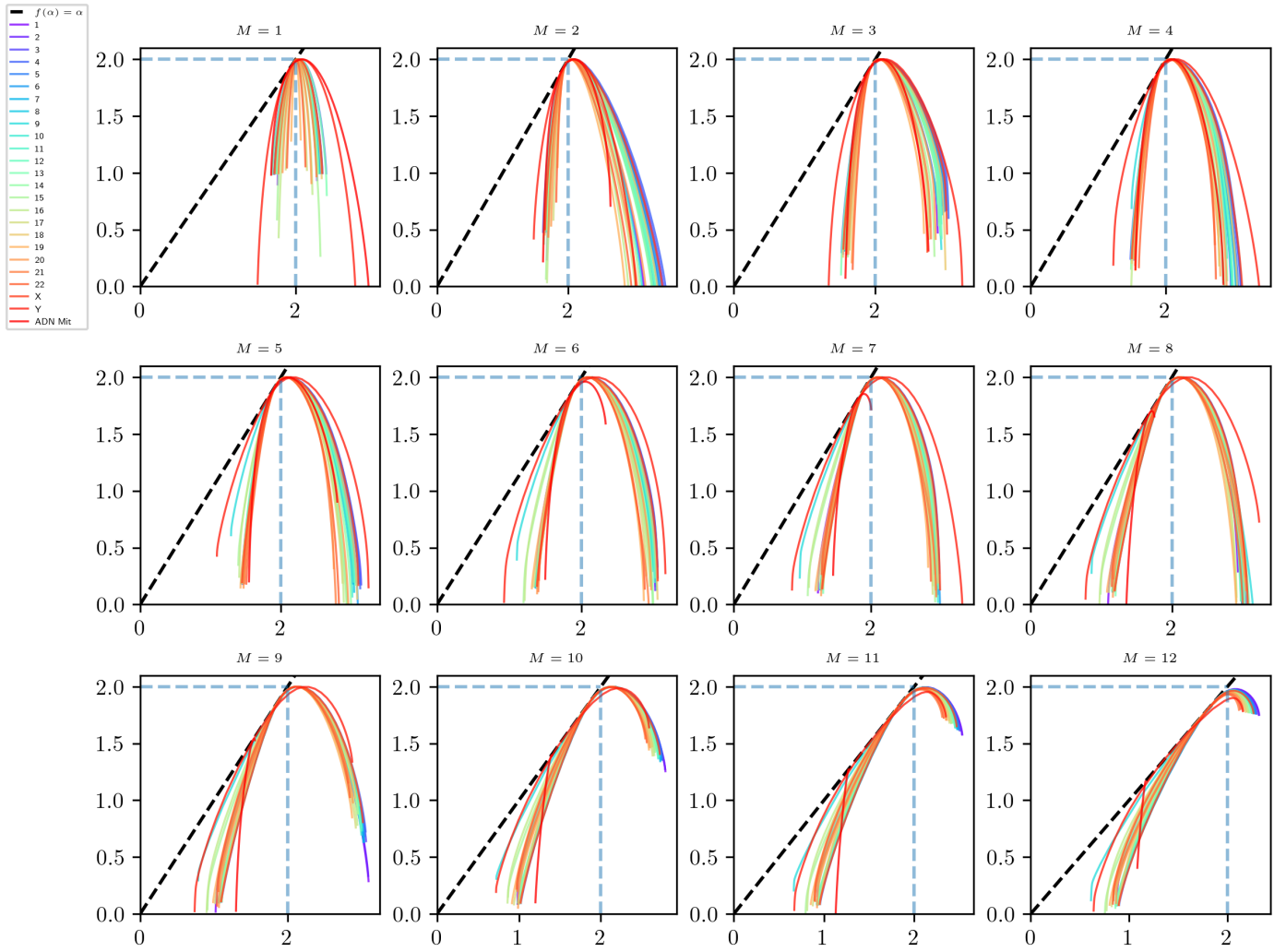
Estas consideraciones permiten mantener las estimaciones de dimensionalidad dentro de un rango compatible con la dimensión topológica del plano. Debe tenerse en cuenta que las representaciones BGR corresponden a conjuntos finitos, con un número máximo de puntos igual a  $4^M$ . Debido a esta finitud, siempre es posible reducir el tamaño de las cubiertas hasta valores arbitrariamente pequeños que contengan cada punto individualmente, lo que conduciría, en el límite, a una dimensión fractal igual a cero. Por esta razón, y con el fin de preservar el carácter aproximado de la estimación multifractal, se mantienen las mismas hipótesis adoptadas en el contexto de las representaciones CGR.

En particular, se supone que las cubiertas no dependen de la distribución específica de los puntos observados ni del número de puntos contenidos en cada región, sino únicamente de la escala geométrica asociada a las posibles secuencias de longitud  $M$ . De este modo, cada cubierta se considera asociada al conjunto de puntos correspondientes a subsecuencias que terminan a la correspondiente secuencia de tamaño  $M$ .

Cabe señalar que, si se desea obtener una estimación más precisa, la escala de las cubiertas podría reducirse adicionalmente hasta un factor de  $(1/2)^L$ , donde  $L$  denota la longitud total de la secuencia analizada. Alternativamente, la medición podría realizarse con respecto a la distancia entre cada punto y los vértices de las cubiertas asociadas a las subsecuencias de longitud  $M$ , lo que permitiría asignar una escala específica a cada subsecuencia observada.

No obstante, la conveniencia de incorporar este nivel adicional de precisión requiere un análisis más detallado. En particular, es necesario evaluar si el refinamiento de las cubiertas produce cambios significativos en las estimaciones multifractales y si tales cambios tienen un impacto relevante en la caracterización práctica de las secuencias. De no ser así, el incremento en la complejidad computacional y metodológica podría no justificarse frente a las ganancias obtenidas en precisión.











# Glosario

**ADN (ácido desoxirribonucleico)** Molécula que transporta la información genética para el desarrollo y funcionamiento de los organismos. Está compuesta por dos cadenas complementarias de nucleótidos que se enrollan formando una doble hélice. Cada cadena tiene una estructura principal de azúcar (desoxirribosa) y fosfato, con bases nitrogenadas (adenina, citosina, guanina, timina) unidas a cada azúcar. La secuencia de estas bases codifica la información biológica. El ADN se organiza en cromosomas dentro del núcleo celular y sus segmentos funcionales se denominan genes..

**ADN satélite** Secuencias de ADN altamente repetitivas, organizadas en largos arreglos en tándem, que constituyen una parte importante de la heterocromatina, especialmente en regiones centroméricas. Su unidad repetitiva puede variar desde motivos muy cortos (menos de 10 nucleótidos) hasta secuencias de más de 100 nucleótidos..

**atractor** Punto fijo del operador de Hutchinson. Es el único conjunto compacto  $A \in \mathcal{H}(\mathbf{X})$  que satisface  $A = W(A) = \bigcup_{n=1}^N w_n(A)$ , donde  $\{w_n\}$  son las contracciones de un IFS. Se obtiene como límite  $A = \lim_{n \rightarrow \infty} W^{on}(B)$  para cualquier  $B \in \mathcal{H}(\mathbf{X})$ ..

**atractor extraño** Conjunto invariante hacia el cual convergen las órbitas de un sistema dinámico disipativo, que exhibe sensibilidad a las condiciones iniciales y posee estructura fractal. A diferencia de los atractores clásicos (puntos fijos, ciclos límite), no es una variedad diferenciable y suele ser autosimilar. Ejemplo: el atractor del mapa logístico en el régimen caótico..

**autoafinidad** Propiedad de un objeto o función que es invariante bajo transformaciones afines que actúan con diferentes factores de escala en distintas direcciones. A diferencia de la autosimilitud, donde la escala es la misma en todos los ejes (transformación de semejanza), en la autoafinidad las contracciones o dilataciones son independientes por dimensión. Esta propiedad es característica de ciertos fractales y funciones, como la función de Cantor (escalera del diablo), cuya gráfica se replica con escalas horizontales y verticales distintas. En el contexto de procesos estocásticos, un movimiento browniano es autoafín, pues sus incrementos escalan de forma diferente en tiempo y en amplitud..

**autosimilitud** Propiedad de un objeto geométrico que puede descomponerse en partes que son copias exactas o aproximadas del todo, pero a una escala reducida. En la autosimilitud exacta, característica de muchos fractales deterministas como el triángulo de Sierpinski o el conjunto de Cantor, las partes se obtienen mediante transformaciones de semejanza que aplican el mismo factor de escala en todas las direcciones, acompañadas de traslaciones y eventualmente rotaciones. En la autosimilitud estadística, propia de fenómenos naturales y multifractales, las distribuciones de probabilidad o los patrones promedio se repiten al cambiar de escala, aunque las configuraciones individuales puedan diferir..

**autosimilitud estadística** Forma débil de autosimilitud en la que las propiedades estadísticas de un objeto o proceso (como momentos, correlaciones o distribuciones de probabilidad) se conservan bajo cambios de escala, aunque las realizaciones individuales no sean exactamente iguales. Es una propiedad central en la teoría de multifractales y aparece en fenómenos como la turbulencia, las series temporales financieras y la organización de secuencias genómicas. A diferencia de la autosimilitud exacta, la autosimilitud estadística no requiere que las partes sean copias geométricas del todo, sino que sus leyes de probabilidad sean invariantes de escala..

**base nitrogenada** Cada uno de los cuatro compuestos orgánicos que forman parte de los nucleótidos del ADN: adenina (A), citosina (C), guanina (G) y timina (T). En el ARN, la timina es reemplazada por uracilo (U). En la Representación del Juego del Caos, cada base se asocia con un vértice del cuadrado unitario..

**bipartición (proceso de)** Proceso iterativo que divide un intervalo en dos mitades iguales en cada paso, asignando pesos probabilísticos  $p_0$  y  $p_1$  a cada subintervalo. Genera una medida fractal sobre  $[0, 1]$  cuya función de distribución acumulada es autoafín y conduce a un espectro multifractal dado por la entropía binaria..

**bola cerrada** Conjunto  $\mathbf{B}(x, \epsilon) = \{y \in \mathbf{X} : d(x, y) \leq \epsilon\}$  que contiene todos los puntos cuya distancia a  $x$  es menor o igual que  $\epsilon$ ..

**cadena de Markov** Proceso estocástico en tiempo discreto con un conjunto finito de estados  $\{1, \dots, N\}$  y una matriz de transición  $(p_{ij})$  donde  $p_{ij}$  es la probabilidad de pasar del estado  $i$  al estado  $j$  en un paso. En los RIFS, la cadena de Markov gobierna la selección de las transformaciones..

**caminante determinista (DW)** En el contexto de CGR, es la trayectoria generada por la aplicación iterativa de las funciones contractivas del IFS en el orden exacto impuesto por la secuencia genómica. A diferencia del juego del caos estocástico, el caminante determinista produce puntos únicos y perfectamente determinados por la secuencia..

**caminante determinista compactado (CDW)** Variante del caminante determinista en la que cada punto codifica un bloque de  $M$  nucleótidos en lugar de uno solo. Esto se logra agrupando la secuencia en bloques de longitud fija y aplicando la transformación correspondiente. Permite representar genomas completos con un número reducido de puntos y superar las limitaciones de precisión numérica..

**caos** Sistema dinámico  $\{X, f\}$  que cumple: (1) es transitivo, (2) es sensible a condiciones iniciales, y (3) el conjunto de órbitas periódicas es denso en  $X$ ..

**cascada multifractal** Modelo iterativo de transferencia de energía en turbulencia, inspirado en la cascada de Richardson. En cada paso, un remolino se subdivide y reparte su energía según proporciones aleatorias fijas, generando una distribución espacial intermitente y multifractal de la disipación. Es análogo a los procesos de bipartición con pesos..

**centrómero** Región especializada y constreñida de un cromosoma que une los dos brazos (brazo p y brazo q). El centrómero juega un papel crucial durante la división celular (mitosis y meiosis), ya que es el punto de unión para las fibras del huso cromático que separan las cromátidas hermanas hacia las células hijas. La posición del centrómero (metacéntrica, submetacéntrica o acrocéntrica) es una característica clave para la identificación cromosómica..

**cilindros (en teoría de medidas)** En la construcción de medidas sobre espacios de sucesiones, los cilindros de orden  $n$  son los subconjuntos correspondientes a sucesiones finitas de longitud  $n$ . Para una medida multinomial en  $[0, 1]$ , los cilindros son los intervalos de longitud  $b^{-n}$  etiquetados por una palabra de  $n$  dígitos en base  $b$ ..

**clausura de un conjunto** Denotada  $\bar{S} = S \cup \{\text{puntos límite de } S\}$ . Es el conjunto cerrado más pequeño que contiene a  $S$ ..

**codificación por bloques** Método de compresión de una secuencia genómica basado en BGR: la secuencia extendida (incluyendo el pseudo-nucleótido inicial) se divide en bloques de longitud fija  $M$  (excepto el último). Cada bloque se codifica como un vector entero  $\vec{N}_\alpha$  mediante conversión binaria a decimal de sus coordenadas. Toda la secuencia queda representada por unos pocos vectores enteros y los parámetros  $N + 1$  y  $M$ ..

**codón** Secuencia de tres nucleótidos en el ADN o ARN que constituye la unidad básica de información genética para la síntesis de proteínas. Cada codón codifica un aminoácido específico o una señal de inicio/terminación de la traducción. Por ejemplo, el codón ATG (o AUG en ARN) codifica metionina y actúa como señal de inicio, mientras que TAA, TAG y TGA son codones de terminación..

**compacto** Subconjunto  $S$  de un espacio métrico tal que toda sucesión infinita en  $S$  contiene una subsucesión que converge a un punto en  $S$ . En  $\mathbb{R}^m$ , equivale a ser cerrado y acotado..

**conexo** Conjunto  $S$  de un espacio métrico tal que los únicos subconjuntos abiertos y cerrados en  $(S, d)$  son  $S$  y  $\emptyset$ . Intuitivamente, es un conjunto de una sola pieza..

**conjugación de densidades** Relación entre las densidades de probabilidad de dos sistemas dinámicos topológicamente conjugados. Si  $\theta$  es la conjugación, entonces las densidades  $\varrho_x$  y  $\varrho_y$  se relacionan mediante  $\varrho_y(y) = \varrho_x(\theta(y)) |\det D\theta(y)|$ , reflejando la conservación de la probabilidad..

**conjugación topológica** Relación entre dos sistemas dinámicos  $\{X_1, f_1\}$  y  $\{X_2, f_2\}$  cuando existe un homeomorfismo  $\theta : X_1 \rightarrow X_2$  tal que  $\theta \circ f_1 = f_2 \circ \theta$ . Los sistemas conjugados tienen propiedades dinámicas equivalentes..

**conjunto de Cantor** Fractal clásico obtenido como atractor del IFS  $\{\mathbb{R}; w_1(x) = x/3, w_2(x) = x/3 + 2/3\}$ . Es un conjunto totalmente desconectado, perfecto y de medida nula, con dimensión fractal  $\ln 2 / \ln 3$ ..

**conjunto de Cantor (medida asociada)** Medida de probabilidad definida sobre el conjunto de Cantor mediante la asignación de pesos  $p_0$  y  $p_2$  a los subintervalos izquierdo y derecho en cada etapa de la construcción ternaria, con  $p_1 = 0$ . Es un ejemplo paradigmático de medida multifractal, con espectro  $f(\alpha)$  dado por una expresión analítica..

**conjuntos de frecuencia idéntica** Agrupaciones de puntos en la BGR que comparten el mismo valor de frecuencia (número de apariciones de la subsecuencia correspondiente en el genoma). Estos conjuntos varían en tamaño y distribución espacial; los de baja frecuencia suelen ser grandes y dispersos, mientras que los de alta frecuencia son pequeños y concentrados. Su estudio revela la estructura repetitiva del genoma..

**continua** Función  $f : \mathbf{X}_1 \rightarrow \mathbf{X}_2$  entre espacios métricos tal que para cada  $\epsilon > 0$  y  $x \in \mathbf{X}_1$  existe  $\delta > 0$  con  $d_1(x, y) < \delta \Rightarrow d_2(f(x), f(y)) < \epsilon$ ..

**cromosoma** Estructura organizada de ADN y proteínas (histonas) que contiene la información genética. En humanos, las células somáticas tienen 46 cromosomas: 22 pares de autosomas y un par de cromosomas sexuales (XX en mujeres, XY en hombres). Cada cromosoma está formado por un brazo corto (brazo p) y un brazo largo (brazo q) unidos por una región constreñida llamada centrómero. La posición del centrómero es una característica distintiva para la clasificación cromosómica..

**cubrimiento óptimo (en CGR)** Familia de cubrimientos del cuadrado unitario obtenida aplicando el operador de Hutchinson  $W$  un número  $n$  de veces. Cada cubrimiento  $C_n$  está formado por  $4^n$  cuadrados de lado  $2^{-n}$ , que corresponden a todas las subsecuencias de longitud  $n$ . Permite localizar exactamente los puntos de la órbita del caminante determinista..

**decodificación BGR** Proceso inverso a la codificación por bloques: a partir de los vectores enteros  $\{\vec{N}_\alpha\}$  y los parámetros  $(N+1, M)$ , se reconstruye la secuencia genómica original sin pérdida de información. Cada vector se convierte a binario y se descompone en los nucleótidos del bloque correspondiente mediante la asignación de vértices..

**denso** Subconjunto  $B \subset X$  cuya clausura es todo  $X$ . En particular, una órbita  $\{x_n\}$  es densa si para cada  $a \in X$  existe una subsucesión convergente a  $a$ ..

**descomposición modal dinámica (DMD)** Técnica numérica que, a partir de datos de una trayectoria, aproxima el operador de Koopman y descompone la dinámica en modos espaciales con frecuencias características. Se utiliza para identificar estructuras coherentes en sistemas dinámicos..

**dimensión de caja** Caso particular de la dimensión de Rényi para  $q = 0$ , denotada  $D_0$ . Coincide con la dimensión fractal del soporte de la medida, calculada mediante el recubrimiento con cajas de lado  $\delta$ :  $D_0 = \lim_{\delta \rightarrow 0} \log N(\delta) / \log(1/\delta)$ , donde  $N(\delta)$  es el número de cajas que intersectan el soporte..

**dimensión de correlación** Dimensión de Rényi de orden  $q = 2$ , denotada  $D_2$ . Mide la probabilidad de que dos puntos elegidos aleatoriamente según la medida estén separados por una distancia menor que  $\delta$ . Es especialmente útil en análisis de series temporales y sistemas dinámicos por su facilidad de estimación..

**dimensión de Hausdorff-Besicovitch** Número  $D_H$  que generaliza la noción de dimensión fractal. Se define a partir de la medida  $p$ -dimensional de Hausdorff: existe un único  $D_H$  tal que la medida es infinita para  $p < D_H$  y cero para  $p > D_H$ . Siempre  $D_H \leq D$  (dimensión fractal)..

**dimensión de información** Dimensión de Rényi de orden  $q = 1$ , denotada  $D_1$ . Está relacionada con la entropía de Shannon de la medida:  $D_1 = \lim_{\delta \rightarrow 0} H_1(\delta) / \log(1/\delta)$ , donde  $H_1(\delta) = -\sum p_i \log p_i$  sobre una partición de tamaño  $\delta$ . Cuantifica la tasa a la que la medida genera información..

**dimensión de Rényi** Familia de dimensiones generalizadas  $D_q$  parametrizadas por un número real  $q$ . Se define a partir de los momentos de la medida sobre una partición de tamaño  $\delta$ :

$D_q = \lim_{\delta \rightarrow 0} \frac{1}{q-1} \frac{\log \sum_i p_i^q}{\log \delta}$  para  $q \neq 1$ , y  $D_1$  como el límite cuando  $q \rightarrow 1$ . Proporcionan una caracterización global de la heterogeneidad de una medida fractal..

**dimensión fractal** Número  $D = \lim_{\epsilon \rightarrow 0} \frac{\ln \mathcal{N}(A, \epsilon)}{\ln(1/\epsilon)}$ , donde  $\mathcal{N}(A, \epsilon)$  es el mínimo número de bolas de radio  $\epsilon$  necesarias para cubrir  $A$ . Mide la complejidad geométrica de un conjunto..

**dirección** Para un IFS con atractor  $A$  y espacio de direcciones  $\Sigma$ , una dirección de  $a \in A$  es cualquier  $\omega \in \Sigma$  tal que  $\phi(\omega) = a$ , donde  $\phi : \Sigma \rightarrow A$  es la función continua y sobreyectiva..

**distancia de Hausdorff** Métrica en el espacio  $\mathcal{H}(\mathbf{X})$  definida como  $h(A, B) = \max\{D(A, B), D(B, A)\}$ , donde  $D(A, B) = \max_{x \in A} d(x, B)$ . Mide la similitud entre dos conjuntos compactos..

**dualidad de Legendre** Relación matemática entre el espectro multifractal  $f(\alpha)$  y el exponente de correlación  $\tau(q)$  mediante la transformada de Legendre:  $\tau(q) = \inf_{\alpha} [q\alpha - f(\alpha)]$  y  $f(\alpha) = \inf_q [q\alpha - \tau(q)]$ . Esta dualidad refleja la equivalencia entre las descripciones local (vía exponentes de Hölder) y global (vía momentos) de una medida multifractal..

**eigenfunción de Koopman** Función  $\phi$  no nula tal que  $U\phi = \lambda\phi$ , donde  $\lambda$  es el valor propio correspondiente. Las eigenfunciones permiten descomponer la dinámica en modos coherentes..

**entropía de Rényi** Generalización de la entropía de Shannon para una distribución de probabilidad discreta  $\{p_i\}$ , definida como  $H_q = \frac{1}{1-q} \log \sum_i p_i^q$  para  $q \neq 1$ , y  $H_1 = -\sum_i p_i \log p_i$  (entropía de Shannon). En el contexto multifractal, la dimensión de Rényi  $D_q$  se obtiene como el cociente  $H_q / \log(1/\delta)$  en el límite de resolución fina..

**ergodicidad** Propiedad de un sistema dinámico con medida invariante tal que los promedios temporales coinciden con los promedios espaciales para funciones observables. Equivalentemente, el único subconjunto invariante no trivial tiene medida cero o total..

**espacio BGR** Conjunto de puntos de la forma  $(i, j)$  con  $i, j$  enteros en el rango  $[0, 2^M - 1]$ , donde cada punto representa una subsecuencia de longitud  $M$  mediante sus coordenadas binarias. La frecuencia de cada subsecuencia se contabiliza en una matriz de  $2^M \times 2^M$ . Es un subconjunto del espacio CGR continuo y permite visualizaciones basadas en densidad..

**espacio CGR** Conjunto de todos los puntos que pueden generarse mediante CGR para una longitud máxima  $N$ . Está formado por la unión de mallas centradas en subcuadrados de lado  $2^{-n}$  para  $n = 1, \dots, N$ . En el límite  $N \rightarrow \infty$  llena densamente el cuadrado unitario. Es el soporte continuo de las medidas asociadas a secuencias genómicas..

**espacio completo** Espacio métrico  $(\mathbf{X}, d)$  donde toda sucesión de Cauchy converge a un punto de  $\mathbf{X}$ ..

**espacio de direcciones** Para un IFS de  $N$  transformaciones, es el conjunto  $\Sigma$  de todas las sucesiones infinitas  $\sigma = \sigma_1\sigma_2\sigma_3\dots$  con  $\sigma_i \in \{1, \dots, N\}$ , dotado de la métrica  $d_c(\omega, \sigma) = \sum_{n=1}^{\infty} |\omega_n - \sigma_n|/(N+1)^n$ .

**espacio de fractales** Espacio  $\mathcal{H}(\mathbf{X})$  cuyos puntos son los subconjuntos compactos no vacíos de  $\mathbf{X}$ . Dotado de la distancia de Hausdorff  $h$ ,  $(\mathcal{H}(\mathbf{X}), h)$  es un espacio métrico completo..

**espacio métrico** Par  $(\mathbf{X}, d)$  donde  $\mathbf{X}$  es un conjunto y  $d : \mathbf{X} \times \mathbf{X} \rightarrow \mathbb{R}$  es una función (métrica) que satisface: (1) simetría, (2) positividad, (3) identidad de indiscernibles y (4) desigualdad triangular..

**espectro multifractal**  $f(\alpha)$  Función que asigna a cada exponente de singularidad  $\alpha$  la dimensión de Hausdorff del conjunto de puntos donde la medida local escala con ese exponente. Formalmente,  $f(\alpha) = \dim_H\{x : \lim_{\delta \rightarrow 0} \log \mu(B_\delta(x))/\log \delta = \alpha\}$ . Describe la distribución de las singularidades de la medida..

**espectro óptico** Representación visual de la estructura multifractal de una medida, en la que diferentes conjuntos  $M(\phi)$  (puntos con frecuencias de dígitos dadas) se colorean según su dimensión fractal, y se superponen para formar un patrón de bandas. Revela la autosimilitud estadística y la distribución de dimensiones en el soporte..

**estructura fractal finita** Familia contable de cubrimientos  $\{C_n\}$  del atractor de un IFS, donde  $C_n = W^n(F)$ . En el contexto CGR, estos cubrimientos son las mallas de cuadrados de lado  $2^{-n}$  que permiten localizar los puntos generados por una secuencia genómica de longitud finita..

**eventualmente periódico** Punto  $x$  de un sistema dinámico tal que  $f^{om}(x)$  es periódico para algún entero positivo  $m$ ..

**exponente de correlación**  $\tau(q)$  Función definida a partir de los momentos de la medida:  $\tau(q) = \lim_{\delta \rightarrow 0} \log \sum_i p_i^q / \log \delta$ . Para una medida multinomial,  $\tau(q) = -\log_b \sum_\omega p_\omega^q$ . Está relacionada con la dimensión de Rényi mediante  $\tau(q) = (q-1)D_q$ . Es la transformada de Legendre del espectro  $f(\alpha)$ ..

**exponente de Hölder** Número real  $\alpha(x)$  que caracteriza el comportamiento local de una medida  $\mu$  en un punto  $x$ , definido por  $\alpha(x) = \lim_{\delta \rightarrow 0} \log \mu(B_\delta(x))/\log \delta$ , cuando el límite existe. Mide la concentración de la medida: valores altos indican regiones de baja densidad, valores bajos indican regiones de alta concentración..

**frecuencia de subsecuencias** Número de veces que una subsecuencia de longitud  $M$  aparece en una secuencia genómica. En BGR, estas frecuencias se almacenan en una matriz de tamaño

$2^M \times 2^M$  y pueden visualizarse mediante escalas de color. Son la base para análisis estadísticos y multifractales..

**frontera de un conjunto** Conjunto  $\partial S$  de puntos  $x \in \mathbf{X}$  tales que toda bola  $\mathbf{B}(x, \epsilon)$  contiene puntos de  $S$  y de  $\mathbf{X} \setminus S$ ..

**función característica (indicatriz)** Función  $\mathbf{1}_A(x)$  que vale 1 si  $x \in A$  y 0 en caso contrario. Se utiliza para definir funciones simples..

**función de Cantor (escalera del diablo)** Función de distribución acumulada de la medida de Cantor. Es continua, monótona no decreciente, con derivada cero en casi todo punto, pero aumenta de 0 a 1 sobre el conjunto de Cantor. Presenta una estructura autoafín y es un ejemplo clásico de función singular..

**función de partición (multifractal)** Análoga a la función de partición en mecánica estadística, definida como  $Z_q(\delta) = \sum_i p_i^q$  para una partición de tamaño  $\delta$ . Su comportamiento de escala  $\log Z_q(\delta) \sim \tau(q) \log \delta$  define el exponente  $\tau(q)$ . Agrupa la información de los momentos de la medida..

**función simple** Combinación lineal finita de funciones características de conjuntos de Borel disjuntos. Se emplea en la construcción de la integral de Lebesgue..

**gen** Unidad física y funcional básica de la herencia. Los genes están formados por secuencias de ADN y actúan como instrucciones para producir moléculas funcionales, principalmente proteínas, aunque muchos genes no codifican proteínas sino que regulan la expresión de otros genes. En humanos, se estima la existencia de aproximadamente 19,900 a 25,000 genes que codifican proteínas..

**genoma** Conjunto completo de material genético de un organismo, incluyendo tanto los genes codificantes como las regiones no codificantes. En humanos, el genoma está organizado en 23 pares de cromosomas y el ADN mitocondrial, comprendiendo aproximadamente 3.1 mil millones de pares de bases. Versiones mejoradas como T2T-CHM13v2.0 han resuelto regiones previamente incompletas..

**genoma T2T-CHM13v2.0** Ensamblaje completo y sin huecos del genoma humano, publicado por el consorcio Telómero a Telómero (T2T). Incluye secuencias de todos los cromosomas y el ADN mitocondrial, con una longitud total aproximada de 3.1 Gb. Es la secuencia de referencia utilizada en este trabajo para los análisis BGR y multifractales..



**heterocromatina** Región del genoma altamente condensada y transcripcionalmente inactiva, rica en secuencias repetitivas y ADN satélite. Se localiza típicamente en los centrómeros y telómeros, y juega un papel importante en la estabilidad cromosómica. En los análisis BGR, las regiones heterocromáticas muestran patrones de alta frecuencia y baja complejidad..

**homeomorfismo** Función biyectiva y continua entre espacios métricos cuya inversa también es continua. Dos espacios homeomorfos son topológicamente equivalentes..

**IFS** Conjunto finito de aplicaciones contractivas  $\{w_n : \mathbf{X} \rightarrow \mathbf{X}\}_{n=1}^N$  sobre un espacio métrico completo  $(\mathbf{X}, d)$ . Se denota  $\{\mathbf{X}; w_1, \dots, w_N\}$  y su factor de contracción global es  $s = \max\{s_n\}$ ..

**integral respecto a una medida** Generalización de la integral de Riemann. Para funciones simples se define como  $\int \sum a_j \mathbf{1}_{B_j} d\mu = \sum a_j \mu(B_j)$ . Para funciones continuas se aproxima por funciones simples..

**interior de un conjunto** Conjunto  $S^0$  de puntos  $x \in S$  para los que existe  $\epsilon > 0$  con  $\mathbf{B}(x, \epsilon) \subset S$ ..

**intermitencia (en turbulencia)** Fenómeno por el cual la disipación de energía en un flujo turbulento no es homogénea en el espacio, sino que se concentra en regiones pequeñas y muy activas, separadas por zonas de menor actividad. Esta propiedad es característica de los procesos multifractales y se modela mediante cascadas..

**juego del caos** Algoritmo de iteraciones aleatorias para dibujar el atractor de un IFS. Se asigna una probabilidad  $p_i > 0$  a cada transformación  $w_i$ , se elige un punto inicial  $x_0$  y se genera la sucesión  $x_n = w_{\sigma_n}(x_{n-1})$  donde  $\sigma_n$  se selecciona al azar según las probabilidades  $p_i$ . La nube de puntos converge al atractor..

**juego del Caos de Barnsley (BCG)** Algoritmo estocástico para generar fractales mediante un sistema de funciones iteradas con probabilidades. Se elige un punto inicial y en cada paso se aplica aleatoriamente una de las transformaciones contractivas, generando una nube de puntos que converge al atractor. Inspiró la versión determinista de Jeffrey para ADN..

**juego del caos para RIFS** Algoritmo para generar trayectorias que aproximan la medida invariante de un sistema de funciones iteradas recurrentes. Se elige un estado inicial según una distribución y se aplica la transformación correspondiente; el siguiente estado se selecciona según la matriz de transición..

**lagunas en CGR** Regiones del cuadrado unitario que no contienen ningún punto de la representación CGR. Corresponden a subsecuencias de una longitud dada que están completamente

ausentes en la secuencia genómica analizada. Las lagunas exhiben autosimilitud y pueden utilizarse para filtrar secuencias de codones o identificar restricciones biológicas..

**mapa logístico** Sistema dinámico en  $[0, 1]$  dado por  $L_\lambda(x) = \lambda x(1 - x)$  con  $0 < \lambda \leq 4$ . Presenta comportamientos complejos incluyendo bifurcaciones y caos..

**matriz de frecuencias BGR** Matriz de dimensión  $2^M \times 2^M$  cuyas entradas contienen el número de apariciones de cada subsecuencia de longitud  $M$  en el genoma. Las coordenadas de la matriz se obtienen interpretando los  $M$  nucleótidos como un número binario en cada eje. Permite visualizaciones y análisis cuantitativos de la composición de secuencias..

**matriz estocástica** Matriz cuadrada de entradas no negativas donde cada fila suma 1. En los RIFS, la matriz  $(p_{ij})$  define las probabilidades de transición entre los estados (transformaciones)..

**media distorsionada**  $P(q)$  Familia de distribuciones de probabilidad derivadas de una medida multinomial original  $\{p_i\}$  mediante el reescalado  $P_i(q) = p_i^q / \sum_j p_j^q$ . El parámetro  $q$  actúa como un “peso” que enfatiza regiones de alta probabilidad ( $q > 0$ ) o baja probabilidad ( $q < 0$ ). La medida correspondiente se condensa en diferentes conjuntos fractales al variar  $q$ ..

**medida de Borel** Medida definida sobre la  $\sigma$ -álgebra de Borel de un espacio topológico. En el contexto de fractales, se emplean para asignar masas a subconjuntos del atractor..

**medida invariante** Medida de probabilidad  $\mu$  que satisface  $\mu = \sum_i p_i \mu \circ w_i^{-1}$  para un IFS con probabilidades. Es el punto fijo del operador de Markov y describe la distribución asintótica de las órbitas aleatorias..

**medida invariante recurrente** Medida invariante para un sistema de funciones iteradas recurrentes (RIFS). Es el punto fijo del operador de Markov generalizado y se obtiene como límite de las iteraciones del juego del caos con transiciones dependientes del estado..

**medida multinomial** Medida de probabilidad sobre el intervalo unitario  $[0, 1]$  construida mediante un proceso iterativo de división en  $b$  subintervalos iguales, asignando pesos fijos  $p_0, \dots, p_{b-1}$  en cada paso. Es un modelo general que incluye como casos particulares la medida de Cantor ( $b = 3$  con  $p_1 = 0$ ) y la medida de bipartición ( $b = 2$ ). Su estructura es multifractal y su espectro se obtiene mediante el teorema de Eggleston..

**mezcla** Una función  $f : X \rightarrow X$  tiene la propiedad de mezcla si para cualquier par de abiertos  $A, B \subset X$  existe  $N$  tal que  $f^{cn}(A) \cap B \neq \emptyset$  para todo  $n > N$ ..

**multifractal** Objeto geométrico o medida cuya estructura de escalamiento no puede describirse con un único exponente fractal, sino que requiere un espectro continuo de dimensiones para

caracterizar las diferentes singularidades locales. Un multifractal puede entenderse como una superposición de subconjuntos fractales entrelazados, cada uno con su propia dimensión..

**máquina de copiado** Algoritmo para aproximar la medida invariante de un RIFS mediante la combinación ponderada de copias reducidas de imágenes de entrada, utilizando las probabilidades de transición como factores de atenuación..

**nucleótido** Molécula orgánica que constituye la unidad estructural básica de los ácidos nucleicos (ADN y ARN). Un nucleótido está formado por tres componentes: una base nitrogenada (adenina, citosina, guanina, timina o uracilo), un azúcar de cinco carbonos (desoxirribosa en el ADN o ribosa en el ARN) y uno o más grupos fosfato. Los nucleótidos cumplen funciones esenciales en el almacenamiento de información genética, transferencia de energía (ATP) y señalización celular..

**números normales** Números reales en  $[0, 1]$  cuya expansión en una base  $b$  contiene cada dígito con frecuencia asintótica  $1/b$ . En el contexto de la medida multinomial, constituyen el conjunto de máxima dimensión ( $f(\alpha) = 1$ ) y corresponden al exponente de Hölder  $\alpha = -\sum (1/b) \log_b p_i$ ..

**operador de Hutchinson** Aplicación  $W : \mathcal{H}(\mathbf{X}) \rightarrow \mathcal{H}(\mathbf{X})$  definida por  $W(B) = \bigcup_{n=1}^N w_n(B)$  para un IFS  $\{w_1, \dots, w_N\}$ . Es una contracción en  $(\mathcal{H}(\mathbf{X}), h)$  con factor  $s = \max\{s_n\}$ ..

**operador de Koopman** Operador lineal  $U$  que actúa sobre funciones observables  $g$  de un sistema dinámico mediante  $Ug(x) = g(f(x))$ . A pesar de que la dinámica subyacente puede ser no lineal,  $U$  es lineal y su estudio espectral proporciona información sobre el sistema..

**operador de Markov** Para un IFS con probabilidades  $\{w_i, p_i\}$ , el operador  $M$  actúa sobre medidas de probabilidad mediante  $M(\nu) = \sum_i p_i \nu \circ w_i^{-1}$ . Es una contracción en el espacio de medidas y su punto fijo es la medida invariante..

**operador de Perron-Frobenius** Operador lineal  $\mathcal{P}$  que actúa sobre densidades de probabilidad según  $\mathcal{P}\varrho(y) = \sum_{x \in f^{-1}(y)} \varrho(x)/|f'(x)|$ . Su punto fijo es la densidad invariante del sistema dinámico..

**partición de un conjunto** Colección finita de subconjuntos medibles, disjuntos y cuya unión es el conjunto total. Se utiliza para aproximar integrales mediante funciones simples..

**parámetro  $M$  (longitud de bloque)** Número entero que define la longitud de las subsecuencias consideradas en la BGR. Las secuencias se dividen en bloques de  $M$  nucleótidos (excepto el último), y cada bloque se codifica en un punto del espacio BGR. Valores mayores de  $M$  aumentan la resolución pero requieren más memoria..

- perfecto** Conjunto cerrado que es igual al conjunto de sus puntos límite (no tiene puntos aislados)..
- periódico** Punto  $x$  de un sistema dinámico tal que  $f^{on}(x) = x$  para algún  $n > 0$ . El menor  $n$  con esta propiedad es el período mínimo..
- polimorfismo** Variación en la secuencia de ADN que ocurre en al menos el 1 % de la población. Los polimorfismos pueden ser de un solo nucleótido (SNP), inserciones, deleciones o variaciones en el número de repeticiones en tándem. Aunque muchos polimorfismos no tienen efectos fenotípicos, algunos contribuyen a la diversidad genética y pueden asociarse con susceptibilidad a enfermedades o respuesta a fármacos..
- pseudo-nucleótido**  $V_0$  Punto inicial artificial añadido al inicio de la secuencia genómica para facilitar la formulación recursiva. Se define como el centro del cuadrado unitario, con coordenadas  $(0.5, 0.5)$  o, en binario,  $(0.1, 0.1)$ . Su inclusión permite expresar todos los puntos de la órbita como sumas ponderadas de los vectores de nucleótidos..
- regiones ricas en CG** Segmentos del genoma con alta densidad del dinucleótido CG. Suelen asociarse con islas CpG cercanas a promotores de genes y presentar una composición de nucleótidos diferente al resto. En los análisis BGR, los conjuntos de baja frecuencia tienden a ubicarse preferentemente en estas regiones..
- Representación del Juego del Caos (CGR)** Técnica introducida por Jeffrey para visualizar secuencias de ADN mediante un mapeo iterativo: cada nucleótido sucesivo se representa como el punto medio entre el punto anterior y el vértice del cuadrado asociado a dicho nucleótido. El resultado es una nube de puntos con estructura fractal que constituye una “huella digital” del genoma..
- Representación del Juego del Caos Entera (iCGR)** Variante del CGR propuesta por Yin, en la que la secuencia genómica se representa mediante un vector entero  $\vec{Y}_N = \sum_{j=0}^N 2^j \vec{V}_j$ . Este vector se encuentra fuera del cuadrado unitario y codifica toda la información de la secuencia. Es equivalente al último punto del CGR tradicional escalado por  $2^{N+1}$ ..
- Representación Genómica Binaria (BGR)** Extensión de la Representación del Juego del Caos (CGR) que expresa las coordenadas de los puntos en base binaria. Permite almacenar toda la secuencia genómica como un par de números binarios, facilitando la codificación compacta, la reconstrucción exacta de la secuencia y el análisis multifractal. Las coordenadas binarias de un punto reflejan directamente los nucleótidos precedentes..
- RIFS** Generalización de un IFS en la que se incluye una matriz estocástica  $(p_{ij})$  que gobierna la probabilidad de aplicar la transformación  $w_j$  después de haber aplicado  $w_i$ . La cadena de

Markov subyacente debe ser irreducible..

**rotaciones cíclicas (de secuencias)** Desplazamientos circulares de los nucleótidos de una secuencia. Por ejemplo, las rotaciones de ACG son CGA y GAC. En el estudio de secuencias repetitivas, las rotaciones de un motivo pueden aparecer con frecuencias similares, aunque en genomas reales raramente son perfectas debido a polimorfismos..

**secuencia genómica (SG)** Sucesión lineal de nucleótidos (A, C, G, T) que constituye el material genético de un organismo. En este trabajo se analiza el genoma humano completo, incluyendo todos los cromosomas y el ADN mitocondrial..

**sensible** Propiedad de un sistema dinámico por la que existe  $\delta > 0$  tal que para cualquier  $x$  y cualquier bola  $B(x, \epsilon)$ , existe  $y \in B(x, \epsilon)$  y  $n \geq 0$  con  $d(f^{on}(x), f^{on}(y)) > \delta$ . Es una de las condiciones del caos..

**soporte de una medida** Conjunto cerrado de puntos  $x$  tales que toda bola abierta centrada en  $x$  tiene medida positiva. Es el menor cerrado fuera del cual la medida se anula. Para un IFS con probabilidades, el soporte de la medida invariante coincide con el atractor..

**subsecuencia** Fragmento contiguo de una secuencia genómica, de longitud fija  $M$ . En BGR, cada subsecuencia de longitud  $M$  se mapea a un punto único en el espacio de representación, y su frecuencia de aparición se contabiliza en la matriz de frecuencias..

**sucesión de Cauchy** Sucesión  $\{x_n\}$  en un espacio métrico tal que para todo  $\epsilon > 0$  existe  $N$  con  $d(x_n, x_m) < \epsilon$  para todo  $n, m > N$ ..

**telómero** Región de secuencias repetitivas de ADN no codificante ubicada en los extremos de los cromosomas. Su función principal es proteger los extremos cromosómicos para evitar que se desgasten, fusionen o degraden. Cada vez que una célula se divide, los telómeros se acortan ligeramente, actuando como un “reloj molecular” que limita el número de divisiones celulares. En células cancerosas, este acortamiento se previene, permitiendo divisiones ilimitadas..

**teorema de Eggleston** Resultado fundamental en teoría de números y multifractales: la dimensión de Hausdorff del conjunto de puntos en  $[0, 1]$  cuyas expansiones en base  $b$  tienen frecuencias asintóticas  $\phi_0, \dots, \phi_{b-1}$  (con  $\sum \phi_i = 1$ ) está dada por  $-\sum \phi_i \log_b \phi_i$ . Este teorema revela la estructura multifractal del intervalo unitario bajo medidas no uniformes..

**teorema de Elton** Resultado fundamental para IFS con probabilidades: para casi toda órbita generada por el juego del caos, el promedio temporal de cualquier función continua converge a su integral respecto a la medida invariante. Establece la equivalencia entre las frecuencias de visita y la medida invariante..

**teorema del collage para medidas** Análogo al teorema del collage para conjuntos: para cualquier medida  $\nu$  y su imagen bajo el operador de Markov  $M(\nu)$ , la distancia a la medida invariante  $\mu_*$  está acotada por  $d_H(\nu, M(\nu))/(1 - s)$ ..

**teorema del punto fijo de Banach** Teorema fundamental que asegura que toda contracción  $f : \mathbf{X} \rightarrow \mathbf{X}$  en un espacio métrico completo posee un único punto fijo  $x_f$ , y para cualquier  $x \in \mathbf{X}$  la iteración  $f^{on}(x)$  converge a  $x_f$ ..

**totalmente acotado** Conjunto  $S$  tal que para cada  $\epsilon > 0$  existe un conjunto finito  $\{y_1, \dots, y_n\} \subset S$  (una  $\epsilon$ -red) con la propiedad de que todo punto de  $S$  dista menos de  $\epsilon$  de algún  $y_i$ ..

**totalmente desconectado** IFS en el que las imágenes del atractor  $w_i(A)$  son disjuntas dos a dos. Cada punto del atractor tiene una única dirección..

**transformación afín** En  $\mathbb{R}^m$ , transformación de la forma  $w(x) = Mx + t$ , donde  $M$  es una matriz  $m \times m$  y  $t$  un vector de traslación. Es contractiva si los autovalores de  $M$  tienen módulo menor que 1..

**transformación contractiva** Aplicación  $f : \mathbf{X} \rightarrow \mathbf{X}$  para la que existe una constante  $0 \leq s < 1$  (factor de contracción) tal que  $d(f(x), f(y)) \leq s d(x, y)$  para todo  $x, y \in \mathbf{X}$ ..

**transformada de Legendre (en multifractales)** Herramienta matemática que conecta el espectro multifractal  $f(\alpha)$  con el exponente de correlación  $\tau(q)$ . Establece que  $\tau(q) = \inf_{\alpha} [q\alpha - f(\alpha)]$  y recíprocamente  $f(\alpha) = \inf_q [q\alpha - \tau(q)]$ . Esta dualidad es análoga a la relación entre energía libre y entropía en termodinámica, donde  $\alpha$  juega el papel de energía,  $q$  el de temperatura inversa y  $f$  el de entropía..

**transitivo** Sistema  $\{X, f\}$  tal que para todo par de abiertos  $\mathcal{U}, \mathcal{V} \subset X$  existe  $n$  con  $\mathcal{U} \cap f^{on}(\mathcal{V}) \neq \emptyset$ ..

**triángulo de Sierpinski** Fractal clásico obtenido como atractor del IFS de tres contracciones afines en  $\mathbb{R}^2$  con factor  $1/2$ . Tiene dimensión fractal  $\ln 3 / \ln 2$ ..

**variaciones (polimorfismos de un solo nucleótido)** Diferencias de un único nucleótido entre secuencias de la misma longitud. En el contexto de CGR, una variación desplaza la posición del punto resultante de manera predecible. Las regiones repetitivas del genoma suelen contener mezclas de diferentes polimorfismos, dando lugar a conjuntos de alta frecuencia pero con ligeras variaciones..

**vector de medida** Concepto introducido para estudiar configuraciones fractales más complejas; se refiere a una colección de medidas que satisfacen un sistema de ecuaciones de invariancia acopladas, generalizando la noción de medida invariante a múltiples componentes..

**órbita** En un sistema dinámico  $\{X, f\}$ , la órbita de  $x \in X$  es la sucesión  $\{f^{on}(x)\}_{n=0}^{\infty}$ .



Casa abierta al tiempo

UNIVERSIDAD AUTÓNOMA METROPOLITANA

## ACTA DE DISERTACIÓN PÚBLICA

No. 00068

Matrícula: 2173805771

Comportamiento Multifractal  
del Genoma Humano.

En la Ciudad de México, se presentaron a las 14:00 horas del día 1 del mes de julio del año 2026 en la Unidad Iztapalapa de la Universidad Autónoma Metropolitana, los suscritos miembros del jurado:

DR. JOSE LUIS DEL RIO CORREA  
DR. JOSE RUBEN LUEVANO ENRIQUEZ  
DR. ALEJANDRO MUÑOZ DIOSDADO  
DR. LEV GUZMAN VARGAS  
DR. JOSE RAFAEL GODINEZ FERNANDEZ

Bajo la Presidencia del primero y con carácter de Secretario el último, se reunieron a la presentación de la Disertación Pública cuya denominación aparece al margen, para la obtención del grado de:

DOCTOR EN CIENCIAS (FÍSICA)

DE: YULIÁN ANDRÉS ÁLVAREZ BALLESTEROS

y de acuerdo con el artículo 78 fracción IV del Reglamento de Estudios Superiores de la Universidad Autónoma Metropolitana, los miembros del jurado resolvieron:

YULIÁN ANDRÉS ÁLVAREZ BALLESTEROS

ALUMNO

REVISÓ

MTRA. ROSALIA SERRANO DE LA PAZ  
DIRECTORA DE SISTEMAS ESCOLARES

**APROBAR**

Acto continuo, el presidente del jurado comunicó al interesado el resultado de la evaluación y, en caso aprobatorio, le fue tomada la protesta.

DIRECTOR DE LA DIVISIÓN DE CBI

DR. JUAN MORALES CORONA

PRESIDENTE

DR. JOSE LUIS DEL RIO CORREA

VOCAL

DR. JOSE RUBEN LUEVANO ENRIQUEZ

VOCAL

DR. ALEJANDRO MUÑOZ DIOSDADO

VOCAL

DR. LEV GUZMAN VARGAS

SECRETARIO

DR. JOSE RAFAEL GODINEZ FERNANDEZ