



UNIVERSIDAD NACIONAL AUTONOMA  
DE MEXICO

FACULTAD DE CIENCIAS

Leyes de potencias en secuencias genéticas y en  
la dinámica simbólica de mapeos unimodales

**T E S I S**

Que para obtener el título de

**F í s i c o**

Presenta

**Manuel Beltrán del Río García**

Director de tesis

Dr. Germinal Cocho Gil



2005



m344174



Universidad Nacional  
Autónoma de México

Dirección General de Bibliotecas de la UNAM

**Biblioteca Central**



**UNAM – Dirección General de Bibliotecas**  
**Tesis Digitales**  
**Restricciones de uso**

**DERECHOS RESERVADOS ©**  
**PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL**

Todo el material contenido en esta tesis esta protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor.



UNIVERSIDAD NACIONAL  
AUTÓNOMA DE  
MÉXICO

**ACT. MAURICIO AGUILAR GONZÁLEZ**  
**Jefe de la División de Estudios Profesionales de la**  
**Facultad de Ciencias**  
**Presente**

Comunicamos a usted que hemos revisado el trabajo escrito:  
"Leyes de Potencias en Secuencias Genéticas y en la dinámica Simbólica  
de Mapeos Unimodales"

realizado por Beltrán del Río García Manuel

con número de cuenta 9954259-9 , quien cubrió los créditos de la carrera de: Física

Dicho trabajo cuenta con nuestro voto aprobatorio.

Atentamente

Director de Tesis

Propietario Dr. Germinal Cocho Gil

Propietario Dr. Francois Allain Leyvraz Waltz

Propietario Dr. Gustavo Carlos Martínez Mekler

Suplente Dr. Denis Pierre Boyer

Suplente Dr. Pedro Eduardo Miramontes Vidal

Consejo Departamental de Física

*Alicia Zarzosa Pérez*  
M. EN C. ALICIA ZARZOSA PEREZ



FACULTAD DE CIENCIAS  
DEPARTAMENTO DE FÍSICA

*G. Cocho*  
*F. Leyvraz*  
*Gustavo Martínez*  
*D. Boyer*  
*P. Miramontes*

Leyes de potencias en secuencias genéticas y en  
la dinámica simbólica de mapeos unimodales

Manuel Beltrán del Río García

Mayo 2005

# Agradecimientos

Agradezco especialmente: ¡A mi papá!. Con mucho cariño a mi familia: a mamá, a mi queridísimo Luis, a mi tío Mario, a mi tía Mague . A mis muchos brazos derechos: a Waco (of course), a Zirahuén, a Ricardo, a Abraham, a mi querido, querido, querido amigo y profesor Samuel Máynez, a Mariana M., a mi buen Aldo, a Herr W.P.K.Z.Z. Epantosamente especial a la Mon. A amigos diversos: al Coche-Luis, a Fernanda Samaniego, a mi primo Polo, a Martina, a Mauricio, al buen Miguel, a Rubén, a Iván, a la señorita Laura, a Martín, a Dagoberto, al Sumo, a Javier, a Cuauhtémoc, a Ulises, a Carlos P. A mis profesores queridos: Al Dr. Lozano, ¡por supuesto a Germi Cocho!, a Gustavo Martínez, a Thomas Seligman, a François Leyvraz, a Pedro Miramontes, a Denis Boyer. A los que agradezco maravillosas clases en la facultad: a R. Tambutti, al Dr. Magaña, a Ana Meda, a Matías Moreno, a Pilar Martínez, a Francisco Marmolejo, al Dr. Boldú, a Rocío Jáuregui, al Dr. Godoy, a Svetlana Dougar-Jabon, a Max y Guillermo Neuman. A larga distancia: a mi querido, querido, "como lo quiero" Lovely Ludwig van Beethoven.

# Indice

<b>1</b>	<b>Manifestación de las Leyes de Potencias</b>	<b>7</b>
1.1	Antecedentes en leyes de potencias . . . . .	7
1.2	Modelización del fenómeno de ley de potencias . . . . .	8
1.3	Densidades invariantes y la distribución Beta de dos parámetros	14
<b>2</b>	<b>ADN Secuenciado y TPBF.</b>	<b>21</b>
2.1	Estructuras generales del ADN y de proteínas . . . . .	21
2.2	TPBF y secuencias de codones . . . . .	27
2.3	TPBF y n-adas en secuencias de ADN . . . . .	31
<b>3</b>	<b>Recursos de la Dinámica Simbólica</b>	<b>34</b>
3.1	Secuencias Simbólicas . . . . .	34
3.2	n-ómeros de la curva logística y TPBF . . . . .	35
3.3	Modelo de expansión modificación . . . . .	38
<b>4</b>	<b>Inversión de Frobenius Perron y TPBF.</b>	<b>44</b>
4.1	Método de Conjugación . . . . .	44
4.2	Problema inverso de FP aplicado a TPBF . . . . .	48
<b>5</b>	<b>Análisis y Resultados</b>	<b>60</b>
5.1	Secuencias Genéticas y Expansión Modificación . . . . .	60
5.2	El mapeo logístico y el modelo de expansión-modificación . . .	62
5.2.1	Intermitencia y bifurcación tangencial inversa . . . . .	62
<b>6</b>	<b>Naturaleza de <i>TPBF</i> (Conclusiones)</b>	<b>70</b>
6.1	La distribución rango-tamaño en secuencias de Codones. . . . .	70

6.2	Distribución rango-tamaño en secuencias binarias del mapeo logístico. . . . .	71
6.3	TPBF como distribución rango-tamaño . . . . .	71
6.4	Modelos generales generativos de TPBF como distribución rango-tamaño. . . . .	72
<b>7</b>	<b>Análisis de Univesalidad de <i>TPBF</i></b>	<b>73</b>
7.1	Series Financieras . . . . .	73
7.2	<i>TPBF</i> en Música Armónica . . . . .	75

# Introducción

La gran diversidad de sistemas que presentan leyes de potencia relacionando sus variables medibles ha inspirado un estudio interdisciplinario desde hace ya más de un siglo, cobrando un auge súbito en las décadas pasadas, cuando modelos de éste estilo fueron encontrados en temas de sociología, neurociencias e ingeniería computacional entre otros.

Muchos de los primeros sistemas en los que se encontraron leyes de potencias (descartando por supuesto, la ley de Gravitación de Newton, la ley de Hooke, la ley de Coulomb, etc. . . ) fueron sistemas de naturaleza social o cultural, como la distribución de tamaños de ciudades en regiones pequeñas o en el número de veces que aparece determinada palabra en un texto cualquiera. Todas estas regularidades fueron notadas a finales del S.XIX y a principios del S.XX, y han tenido desde entonces un continuo debate en cuanto a qué mecanismo produce las leyes de potencia en ellas. Entre los temas actuales de estudio de las leyes de potencia se encuentran los grados de entrada y salida en redes complejas, en donde redes del tipo WWW generadas con algoritmos aleatorios presentan con frecuencia leyes de potencia, en los modelos neuronales de organismos sencillos, se encuentra también que muchas de las señales nerviosas son leyes de potencias y también en varios modelos económicos existe evidencia experimental de leyes de potencia.

Si llevamos a las leyes de potencia un paso más hacia la complejidad obtenemos la función Beta de dos parámetros<sup>1</sup>, que es el producto de dos funciones potenciales (leyes de potencia  $y = cx^\alpha$ ), una con variable  $x$  y la otra con variable  $N - x$ , si tomamos el dominio de la función como  $[0, N]$ . La función Beta de dos parámetros es otra relación que suele aparecer en sistemas de muy diversas naturalezas: en la generación de números pseudoaleatorios, en flujos turbulentos ([10]), en algunos parámetros medibles del lenguaje es-

---

<sup>1</sup>Nos referimos a la ec(1.18) como función Beta de dos parámetros, que es la forma sin integrar de la función Beta común.



crito ([4]) y en la generación de secuencias simbólicas de diferentes tipos, entre otros. En este último ámbito es en donde existe un ejemplo particularmente interesante, que es el de las secuencias simbólicas extraídas del material genético. La doble cadena de símbolos (bases nitrogenadas) del ADN se interpreta naturalmente como una sola cadena de grupos de tres símbolos llamados codones. Si se cuenta la frecuencia de aparición de cada uno de los 64 distintos grupos en toda la cadena y se organizan los valores de frecuencia de aparición de mayor a menor obtenemos invariablemente una distribución Beta, independientemente de la especie o el individuo. Si en vez de contar grupos de tres símbolos lo hacemos con grupos mayores, la distribución Beta sigue apareciendo y lo hace con mejor resolución.

Para estudiar la aparición de la función Beta en la distribución de codones y  $n$ -adas en el código genético utilizamos un modelo de generación de secuencias binarias que imita los mecanismos de mutación más frecuentes en las secuencias genéticas. El modelo (modelo de expansión-modificación) presenta también funciones Beta en la distribución de grupos de  $n$  símbolos consecutivos.

Con el propósito de encontrar mecanismos generales que conduzcan a la función Beta como en las secuencias de codones y en las distribuciones del modelo de expansión modificación, se produjeron secuencias binarias iterando una condición inicial en el mapeo logístico y obteniendo sus símbolos dependiendo de que si la iteración cayó en una mitad u otra del dominio del mapeo. Las secuencias obtenidas se ajustaban a una función Beta dependiendo del valor del parámetro  $\mu$  del mapeo.

En el primer capítulo se hace un recuento más amplio de los sistemas que presentan comportamiento en forma de leyes de potencias, la manera en que se han abordado los problemas para encontrarles un modelo convincente. También se incluyen comentarios acerca de la información que puede proveer una función de densidad de probabilidad acerca del sistema del cual se extrae. Por último se presenta la función Beta de dos parámetros.

En el segundo capítulo se listan los resultados de ajustar una función Beta de dos parámetros a la distribución de codones y otros grupos de secuencias genéticas, además de dar una introducción elemental a la estructura del ADN y sus principales mecanismos de mutación.

El tercer capítulo trata las características de los dos modelos generadores de secuencias binarias que utilizamos, el modelo de expansión-modificación y las iteraciones del mapeo logístico, así como los ajustes de las distribuciones que producen estos modelos al compararlos con la función Beta.

En el cuarto capítulo se analiza la función Beta creando mapeos caóticos completos mediante el método inverso de Frobenius-Perron, de manera que sus densidades de probabilidad resulten ser (por construcción) funciones Beta de dos parámetros.

La comparación de los mecanismos responsables de la aparición de la función Beta en el modelo de expansión modificación, en el mapeo logístico y en las secuencias de n-ómeros de pares de bases en ADN se hace en el quinto capítulo, a manera de discusión se advierten las semejanzas entre los mecanismos generativos.

Las conclusiones se dan en el sexto capítulo, y en el séptimo presentamos una discusión acerca de la generalidad de los modelos generativos discutidos a lo largo de la tesis.

# Capítulo 1

## Manifestación de las Leyes de Potencias

### 1.1 Antecedentes en leyes de potencias

Ya sea como ajustes empíricos o como conjeturas teóricas, las leyes de potencias (i.e. correlaciones del tipo  $y = cx^\alpha$ ) abundan en el estudio de los fenómenos naturales, en algunos sistemas cuantificables de carácter social e inclusive en la parametrización cruda de algunas manifestaciones artísticas ([1, 6]). A manera de primer ejemplo podemos notar que la formulación matemática de una gran cantidad de leyes clásicas de física tienen el tipo de ley de potencia, entre ellas destacan la ley de gravitación de Newton <sup>1</sup>, la fuerza electrostática entre cargas puntuales (Coulomb) y cualquier aproximación a primer orden de dos variables físicas (la fuerza de restauración elástica de Hooke, por ejemplo). A mediados de los años treinta Zipf descubrió una regularidad en forma de ley de potencia en la distribución de poblaciones en las distintas ciudades de una región determinada ([3]). Zipf hizo una lista con todas las ciudades con una población por arriba de un cierto valor en una región de elección arbitraria (un país o una provincia por ejemplo). Ordenó la lista de manera descendente y encontró que una línea recta de pendiente igual a  $-1$  ajustaba los datos de manera satisfactoria en

---

<sup>1</sup>Aunque una propiedad fundamental de las leyes de potencias es la carencia de una escala propia, la validez a grandes distancias de la ley de gravitación universal se ha puesto en duda para ajustar modelos cosmológicos, debemos entonces considerarla una ley de potencias solo en un rango finito.

escalas log-log. Otra contribución importante de Zipf, más conocida y más llamativa, es la distribución de las frecuencias con que aparecen las distintas palabras utilizadas en un texto cualquiera ([2],[1]) lo suficientemente grande para ser considerado una buena muestra). La relación que encontró entre el número de veces que aparece una palabra y su rango (i.e. que posición ocupa en una lista de todas las palabras del texto ordenada por el número de veces que aparece cada una, en donde la primera palabra es aquella que aparece más veces en el texto y la última aquella que está escrita menos veces) es una ley hiperbólica (ley de potencia con  $\alpha = -1$ ,) (ver Figs. 1.1, 1.2 y 1.3). Otras formas de leyes de potencias se dan en las distribuciones rango-tamaño (*rank size distribution*) de los aerolitos que bombardean a la Tierra con respecto a su diámetro, o en las especies animales con respecto a su masa corporal ([1],[2]). Es necesario mencionar también la aparición de regularidades tipo ley de potencia en modelos matemáticos de sistemas económicos.

Establecer la naturaleza de ley de potencias entre los valores empíricos de dos variables es, por supuesto, solo el primer paso en una investigación formal del comportamiento de éstas, el consecuente desarrollo de una teoría formal puede tomar muchos caminos diferentes. Para la mayoría de los ejemplos mencionados anteriormente se ha optado por construir modelos estocásticos con variables aleatorias, a manera de eliminar la necesidad de caracterizar parámetros complicados de manejar, tales como son, tomados de los ejemplos, las condiciones económicas y sociales que rigen el crecimiento (o decrecimiento) de ciudades y la fundación de nuevos asentamientos. En caso del último ejemplo es evidente que el problema se tiene que abordar de alguna manera poco convencional.

## 1.2 Modelización del fenómeno de ley de potencias

Aunque es improbable que exista un único mecanismo atrás del fenómeno de la ley de potencia, es probable que existan modelos generales para grupos representativos de los problemas de las áreas antes mencionadas ([2, 3]). Por ejemplo, consideremos la ecuación diferencial de Itô:

$$dX = \mu X dt + \sigma X dw \quad (1.1)$$

Esta ecuación representa un movimiento Browniano, se puede utilizar para modelar la evolución temporal de precios de acciones, de ganancias indi-

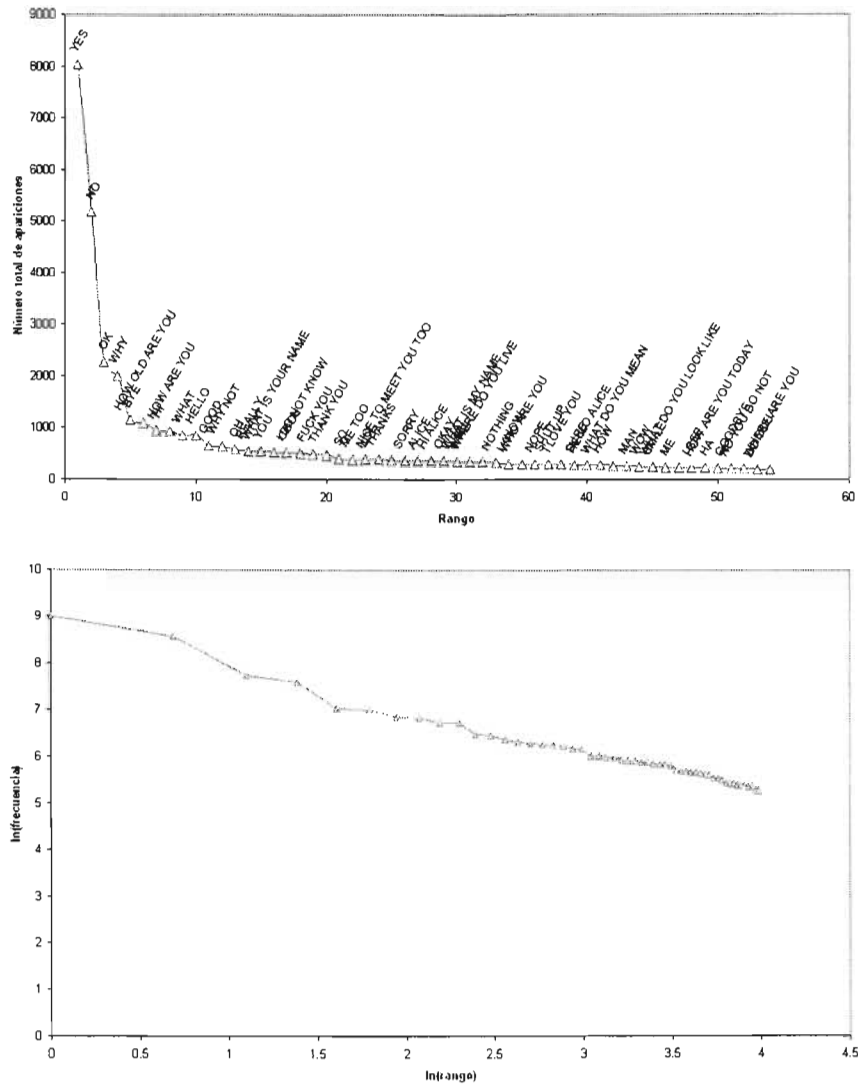


Figura 1.1: Ley de Zipf en lenguajes

## CAPÍTULO 1. MANIFESTACIÓN DE LAS LEYES DE POTENCIAS 10

viduales en cuentas bancarias, etc. En dichos casos  $\mu$  es una componente ajustada empíricamente y  $\sigma$  es un atributo de aleatoriedad que controla un factor de ruido blanco  $dw$ . La distribución que genera éste modelo es una función log-normal. Sin embargo en [3] Reed muestra que si el tiempo de evolución del modelo se toma con una distribución exponencial, entonces las distribuciones rango-tamaño del modelo cambian de log-normales a dobles leyes de potencia:

$$\rho(x) = \begin{cases} \frac{\alpha\beta}{\alpha+\beta} \left(\frac{x}{x_0}\right)^{\beta-1} & x < x_0 \\ \frac{\alpha\beta}{\alpha+\beta} \left(\frac{x}{x_0}\right)^{-\alpha-1} & x \geq x_0 \end{cases} \quad (1.2)$$

Los datos recopilados sobre distribuciones de tamaños de ciudades evidencian la exactitud del modelo en ambas *colas* (en  $x \geq x_0$  y en  $x < x_0$ , y más específicamente: en el grupo de las  $n$  ciudades más pequeñas y en el grupo de las  $m$  ciudades más grandes), y aún más, ésta modificación en la ecuación hace también que el modelo sea aplicable en muchos otros sistemas regidos por leyes de potencias. Si hacemos analogía entre el crecimiento de una ciudad y el crecimiento de el ingreso económico de un individuo y entre el tiempo transcurrido desde la fundación de la ciudad y el de apertura de la cuenta de ahorros, encontramos que la ley de ingresos individuales (o ley de *Pareto*) está respaldada teóricamente por el modelo general de movimiento Browniano con la modificación de Reed (a la que el llama *double Pareto distribution*) y también concuerda con un modelo de caminante aleatorio con barreras reflectoras, en el cuál la probabilidad de un individuo de aumentar o disminuir su capital depende probabilísticamente de la magnitud de éste y en donde también se propone un valor mínimo y uno máximo para el capital, valores que de ser rebasados (hacia abajo o hacia arriba respectivamente) provocarían la restitución del capital en valores permitidos ([16],[17]). De igual manera se ha identificado esta distribución con sistemas de otras áreas distintas de la economía, como en el caso de la distribución de incendios en un bosque, o como en la aparición de distintas especies de un mismo género (*genus*). En el caso de lenguajes, Zipf, como una primera idea para explicar la aparición de la ley hiperbólica, propone un mecanismo de *mínimo esfuerzo*, en donde el desarrollo evolutivo de un lenguaje oblige, de una manera natural, a que el conjunto de palabras siga una evolución hacia la economía. Mandelbrot desarrolla ésta idea explicando que posiblemente la ley de potencia aparezca como una minimización del coeficiente costo/información. La demostración se hace estableciendo el problema de crear un idioma con  $m$

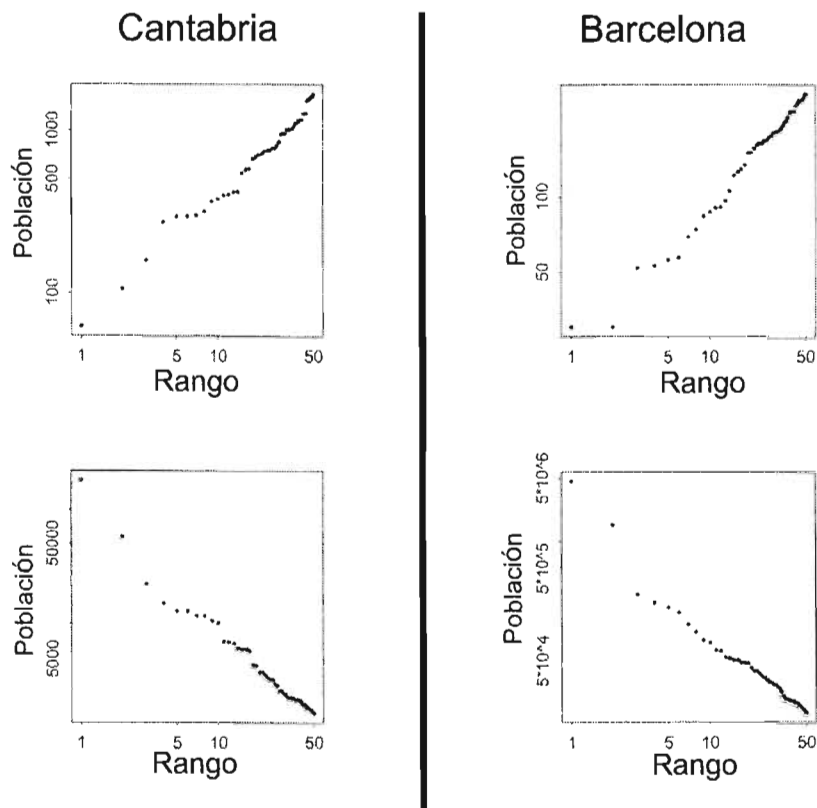


Figura 1.2: datos de ciudades con modelo de Reed

CAPÍTULO 1. MANIFESTACIÓN DE LAS LEYES DE POTENCIAS 12

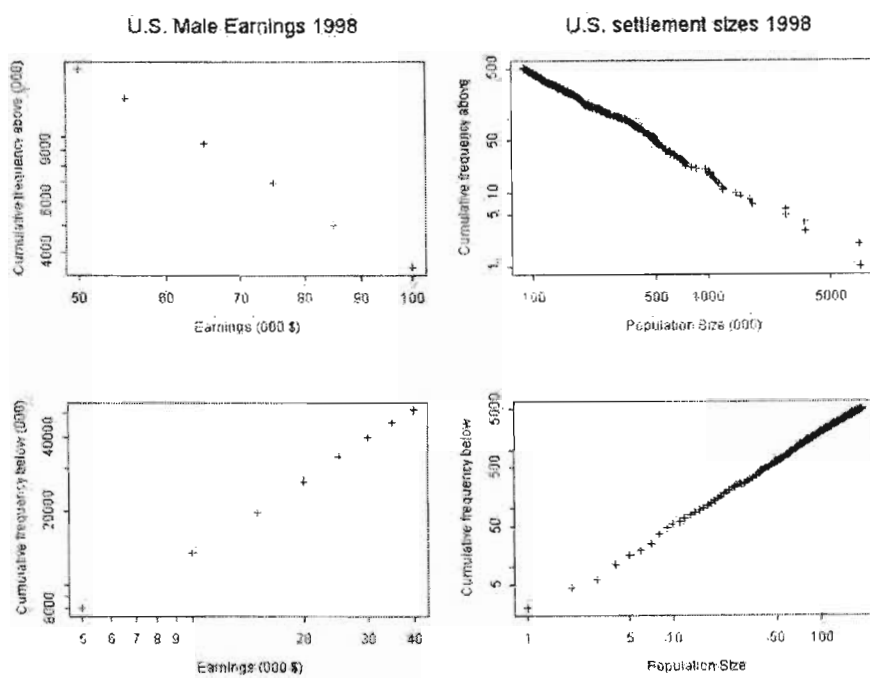


Figura 1.3: ley en ingresos individuales



palabras diferentes constituidas por cadenas fabricadas de un alfabeto de  $n$  letras, el cuál minimice el costo por unidad de información. Propone que el costo por utilizar una palabra sea proporcional al número de letras que utilice, y que el rango de las palabras vaya también como el número de letras que contenga, entonces establece que la función natural de costo para una palabra con rango  $j$  sea  $C_j = \log_n(j)$ , y por lo tanto el costo promedio por palabra está dado por  $C = \sum_{j=1}^m p_j C_j$ , en donde  $p_j$  es la probabilidad de encontrar la palabra de rango  $j$  en un texto cualquiera. La información promedio es tomada como la entropía:  $H = -\sum_{j=1}^m p_j \log_2(p_j)$ . Para minimizar el cociente  $A = \frac{C}{H}$  derivamos:

$$\frac{dA}{dp_j} = \frac{C_j H + C \log_2(p_j e)}{H^2} \quad (1.3)$$

Para anular la derivada obtenemos:

$$p_j = \frac{2^{-\frac{HC_j}{C}}}{e} \quad (1.4)$$

Que tiene forma de ley de potencias dada la forma del precio de las palabras:  $C_j = \log_d(j)$ .

Se demuestra una seria objecion a esta idea <sup>2</sup> alegando que un *idioma* con palabras construidas con una longitud aleatoria, de probabilidad  $p_n = (l_1)^n$  (donde  $n$  es la longitud de la palabra y  $0 < l_1 < 1$  es la probabilidad de que la palabra tenga solo un carácter), en donde cada carácter se escoge de un conjunto finito de manera equiprobable, también sigue una ley hiperbólica.

La demostración se hace de la manera siguiente, consideramos un alfabeto de  $n$  letras diferentes, las palabras se crean escogiendo letras o espacios vacios con probabilidad  $q$  de ser un espacio o probabilidad uniforme  $\frac{1-q}{n}$  de ser cualquier letra. La probabilidad de que aparezca cualquier letra con  $m$  caracteres es entonces  $p_m = (\frac{1-q}{n})^m q$  y como por cada rango existen  $n^m$  palabras, entonces, sustituyendo el exponente  $m$  en la ecuación de probabilidad por  $m \propto \log_n(r)$ , la probabilidad de que aparezca una palabra por su tamaño es:

$$p_m = q \left( \frac{1-q}{n} \right)^{\log_n(r)} = q(r)^{\log_n(1-q)-1} \quad (1.5)$$

, que es una ley de potencias.

---

<sup>2</sup>En [1] cuentan que fue Mandelbrot mismo quien atacaba al modelo de optimización, atribuido a Zipf, mientras que en [16] se sostiene que fue el psicólogo Miller quien dió el argumento de los lenguajes aleatorios en contra de la idea de optimización, de Mandelbrot.

La verdadera naturaleza de esta regularidad sigue siendo tema de discusión, inclusive se han hecho nuevos esfuerzos para explicar la ley de Zipf mediante optimización (economía) de esfuerzos emisor-receptor ([5]). Se ha propuesto, desde un punto de vista mas general, la posibilidad de que exista una estructura del lenguaje propia de la especie ([4]), a través de la cuál se manifiesten restricciones que moldean la forma en que un lenguaje se desarrolla individualmente o en sociedad. La existencia de estas restricciones con carácter universal explicaría las semejanzas estructurales -la ley de Zipf entre ellas- reportadas entre lenguajes con orígenes culturales independientes.

Una solución similar se puede buscar para explicar la constante aparición de leyes de potencia como correlaciones entre las variables físicas de la composición musical, aunque en este caso el *corpus* de estudio tiene una mayor dependencia a su origen cultural [6] que en el caso de lenguajes.

### 1.3 Densidades invariantes y la distribución Beta de dos parámetros

Muchas veces, en el estudio de sistemas físicos, la única información que le es posible extraer al investigador es de carácter estadístico, una forma típica de este conocimiento es la *densidad invariante*. La densidad invariante  $\rho(x)$  es una función que condensa la información estadística acerca de la probabilidad de un sistema de encontrarse en un estado determinado, sin conocimiento de su estado previo. En el ámbito experimental se suele obtener la densidad de manera natural en sistemas en donde la observación directa o detallada es difícil. La densidad invariante es también información útil en el estudio de sistemas que tienen ya un modelo teórico, ya sea por que se ha desarrollado éste de manera específica o por tratarse de un sistema "prefabricado", a éstos últimos corresponde el caso de los mapeos dinámicos.

La función de densidad invariante es particularmente importante en el caso de los mapeos dinámicos dado que las propiedades de ésta primera están fuertemente relacionadas con la *ergodicidad* propia del mapeo, característica importante en el estudio de las propiedades de sistemas caóticos, del cuál son herramienta típica los mapeos dinámicos.

La densidad invariante  $\rho(x)$  se define para mapeos unimodales como:

$$\rho(x) = \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{i=0}^N \delta[x - f^i(x_0)] \quad (1.6)$$

Donde  $x_0$  es una condición inicial arbitraria; si la densidad invariante no depende de dicha condición decimos que el mapeo sobre el cuál esta calculada es *ergódico*, y esta definición es, por supuesto, equivalente a la condición de que exista un subconjunto del dominio con medida igual a la medida del dominio, tal que, cualquier punto inicial en este conjunto estará, después de un número suficientemente grande de iteraciones del mapeo, tan cerca de cualquier otro punto del subconjunto en cuestión como se desee. Cuando éste es el caso, de la anterior definición se puede derivar una manera natural de calcular la densidad invariante utilizando métodos numéricos, discretizando el intervalo unitario y contando el número de iteraciones que caen en cada intervalo después de un número grande de evaluaciones. La posibilidad de que alguna de las iteraciones del mapeo resulte ser un punto fijo inestable es mínima dado que el redondeo efectuado en cada operación ocasionado por la precisión finita funciona como una pequeña perturbación que evita que todos los iterados subsecuentes de un punto fijo inestable permanezcan en la región inmediata a éste.

A veces es posible derivar la densidad invariante de manera analítica a través de la ecuación de *Frobenius-Perron*:

$$\rho(x) = \int_0^1 \delta[x - f(y)]\rho(y)dy \quad (1.7)$$

En donde  $\rho(x)$  se define ahora como el eigenvector de eigenvalor unitario del operador de Frobenius-Perron:

$$\hat{F}_f(A) = \int_0^1 \delta[x - f(y)]A(y)dy \quad (1.8)$$

El mapeo logístico (Fig 1.4):

$$x_{n+1} = f(x_n) = \mu x_n(1 - x_n). \text{ con: } \mu \in [0, 4] \quad (1.9)$$

Es el ejemplo favorito para mostrar de manera clara fenómenos como: la ruta a caos mediante bifurcaciones, pérdida de la historia de los iterados, atractores, puntos fijos estables e inestables, etc. . . y gracias a que se puede resolver para  $\mu = 4$ , es además un ejemplo de un mapeo del cuál puede ser calculada la densidad invariante de manera analítica (para  $\mu = 4$ ): Primero hay que convertir el mapeo logístico de su forma usual:

$$f(x_n) = \underbrace{f(f(f(\dots f(x_0))))}_{n \text{ veces}} \quad (1.10)$$

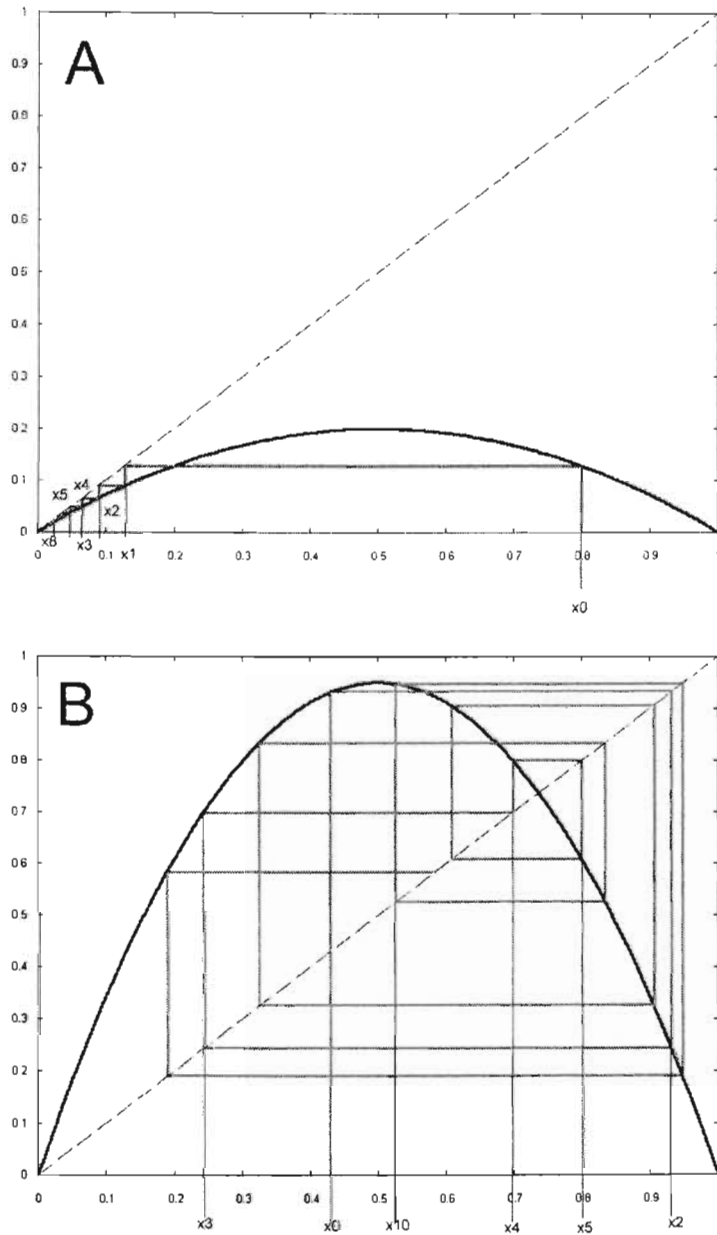


Figura 1.4: mapeo logístico

## CAPÍTULO 1. MANIFESTACIÓN DE LAS LEYES DE POTENCIAS 17

a la forma más práctica:

$$f(x_n) = g(n, x_0). \quad (1.11)$$

Ésto es posible si hacemos el siguiente cambio de variable:

$$x_n = \frac{1}{2}(1 - \cos(2\pi y_n)) = g(y_n) \quad (1.12)$$

entonces la ec.(1.9) se convierte en:

$$\begin{aligned} x_{n+1} &= 4x_n(1 - x_n) \\ \frac{1}{2}(1 - \cos(2\pi y_{n+1})) &= (1 - \cos(2\pi y_n))(1 + \cos(2\pi y_n)) \\ &= 1 - \cos^2(2\pi y_n) \\ &= \frac{1}{2}(1 - \cos(4\pi y_n)). \end{aligned} \quad (1.13)$$

De ésta última podemos concluir que:

$$\begin{aligned} 2\pi(y_{n+1}) &= 4\pi y_n \text{ mod } 2\pi \\ y_{n+1} &= 2y_n \text{ mod } 1 \\ &\equiv f_B. \end{aligned} \quad (1.14)$$

A la relación en forma de mapeo (1.14) se le conoce como mapeo de Bernoulli (*Bernoulli shift*, Fig.1.5), y es un mapeo ergódico con densidad inavriante ( $\rho_B(x)$ ) constante:

$$\rho_B(x) = 1. \quad (1.15)$$

Para acabar de derivar la densidad invariante ( $\rho_L(x)$ ) de (1.9) utilizamos la ec de Frobenius-Perron y hacemos un cambio de variable:

$$\begin{aligned} \rho_L(x) &= \int_0^1 \delta[x - f_L(p)]\rho_L(p)dp \\ &= \int_0^1 \delta[x - g(f_B(p))]\rho_L(p)dp \\ &= \int_0^1 \delta[x - g(z)]\rho_B(z)dz \\ &= \int_0^1 \delta[x - g(z)]dz \end{aligned}$$

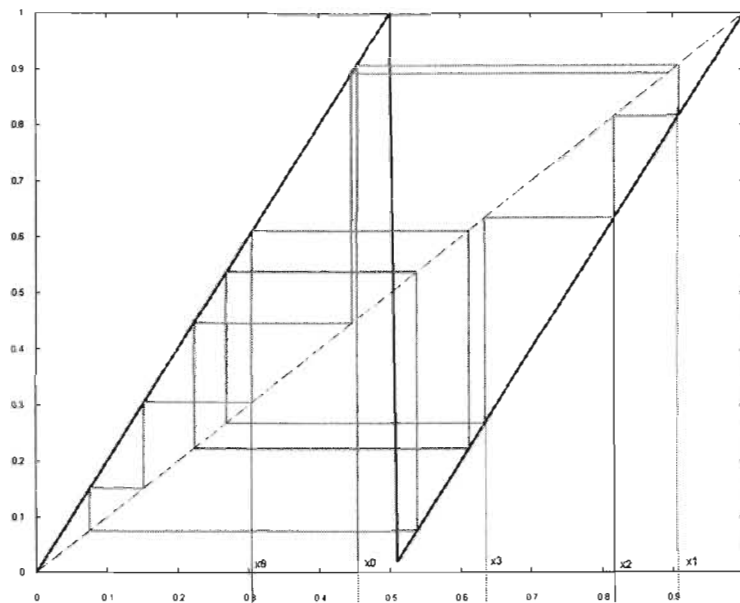


Figura 1.5: Mapeo de Bernoulli

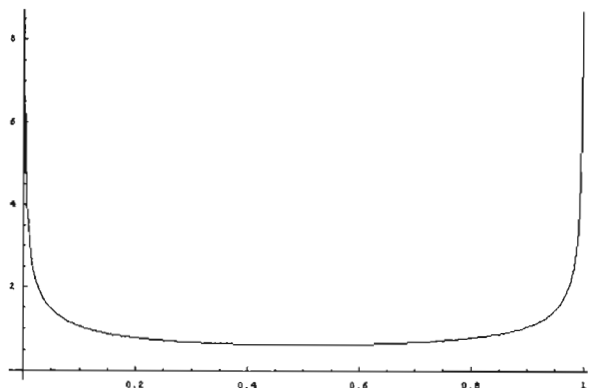


Figura 1.6: densidad del mapeo logístico ( $\mu = 4$ )

finalmente, recordando que:

$$\delta(g(x)) = \sum_a \frac{\delta(a)}{|g'(a)|} \quad (1.16)$$

donde la suma se hace sobre todos los valores  $a$ , en donde la función  $g$  tiene un cero simple ( $g'(a) \neq 0$ ), obtenemos:

$$\begin{aligned} &= \sum_{g_i^{-1}} \int_0^1 \frac{\delta[z - g_i^{-1}(x)]}{|g'(g_i^{-1}(x))|} dz \\ &= \frac{2}{|g'(y)|} \\ \rho_L(x) &= \frac{1}{\pi} \frac{1}{\sqrt{x(1-x)}}. \end{aligned} \quad (1.17)$$

Un resultado importante que se puede derivar de la expresión (1.17) es que, como no depende de una condición inicial arbitraria, el mapeo logístico con  $\mu = 4$  es un mapeo ergódico (Fig. 1.6).

Además de las leyes de potencias, otra forma frecuente que llevan las funciones de densidad es la *distribución Beta de dos parámetros* (**TPBF**):

$$\rho(x) : [0, 1] \longrightarrow \mathfrak{R}$$

$$\rho(x) = \frac{x^\alpha(1-x)^\beta}{B(\alpha+1, \beta+1)} \quad (1.18)$$

en donde  $B(x, y)$  es la función Beta:

$$B(x, y) = \int_0^1 t^{x-1}(1-t)^{y-1} dt \quad (1.19)$$

que aparece aquí como factor de norma.

Densidades con esta forma han aparecido en sistemas de muy diversas naturalezas, tales como: flujos turbulentos, difusión en medios desordenados, *etc ...*<sup>3</sup>, y han sido también utilizadas como funciones "de parecido" (*likelihood function*) para clasificación de imágenes [11].

La distribución (1.18) juega un papel importante en en la relación entre frecuencia y rango en las ternas de bases nitrogenadas (codones) que aparecen en secuencias genéticas de varias especies.

---

<sup>3</sup>Lowe, Frenkel, van der Hoef



# Capítulo 2

## ADN Secuenciado y TPBF.

### 2.1 Estructuras generales del ADN y de proteínas

Los seres vivos deben la mayor parte de su peso a materiales inorgánicos, como son el agua y algunos iones disueltos en ella, pero las propiedades importantes que hacen que un organismo se comporte de la manera ordenada que lo caracteriza se deben a los llamados componentes orgánicos. Estos componentes se clasifican en dos grupos:

**Moléculas pequeñas** -Son los productos de las reacciones metabólicas que procuran energía y demás materiales a la célula.

**Macromoléculas** -Estas son estructuras más complicadas ensambladas todas a partir de moléculas pequeñas.

Dado que las macromoléculas son solamente agregados de moléculas pequeñas podemos decir que todas las estructuras orgánicas se conforman de moléculas pequeñas. Existen cuatro tipos principales de moléculas orgánicas pequeñas:

1. Ácidos grasos-Se forman al enlazar un ión negativo (generalmente un hidrocarburo ionizado) con un grupo carboxilo  $COOH$ .
2. Azúcares-Moléculas del tipo  $C_aH_bO_c$ , entre las cuales están, por ejemplo, la glucosa ( $C_6H_{12}O_6$ ) y la sucrosa ( $C_{12}H_{22}O_{11}$ )
3. Aminoácidos-Tienen la forma general que se muestra en la figura 2.1: en donde R es una molécula llamada *grupo lateral*, que caracteriza a

cada uno de los veinte aminoácidos naturales conocidos (ver también fig 2.1).

4. Nucleótidos-Consisten en un anillo de carbonos y nitrógenos, un azúcar de cinco carbonos también en forma de anillo y un grupo fosfato. Existen solo cinco diferentes: Adenina, Guanina, Timina, Citosina y Uracilo, pero en el ADN solo aparecen cuatro (no hay Uracilo), al igual que en el RNA (en donde el Uracilo sustituye a la Timina del ADN. Ver Fig 2.3).

Las macromoléculas se forman mediante la unión de moléculas pequeñas de un mismo tipo, es entonces que los ácidos grasos forman lípidos, los azúcares forman polisacáridos, los nucleótidos forman ADN y los aminoácidos se unen para formar péptidos. Los aminoácidos se unen mediante un enlace peptídico a través de la condensación del grupo amina ( $NH_2$ ) de un aminoácido con el grupo carboxilo de otro aminoácido, liberando entonces una molécula de agua. Estos enlaces se pueden repetir indefinidamente para formar una cadena de aminoácidos, con un grupo amina en un extremo (extremo "N") y un grupo carboxilo en el otro (extremo "C") y cuya columna central está formada por repeticiones del grupo general de los aminoácidos ( $NH_2 - CH - COOH$ ) y a cada uno de estos eslabones está enlazado el grupo lateral propio del aminoácido (ver Fig 2.2).

A una cadena grande de péptidos se le llama polipéptido. Los polipéptidos a su vez se unen para formar unidades funcionales llamadas proteínas. A las proteínas se les puede caracterizar casi de manera completa, dada su estructura lineal, con la secuencia de aminoácidos que la forman, desde el extremo "N" hasta el extremo "C". A la secuencia de aminoácidos que caracteriza a una proteína se le conoce como estructura primaria. A la manera en que esta cadena enorme de aminoácidos se pliega en un espacio tridimensional se le llama estructura secundaria, ejemplos típicos de estructura secundaria son las hélices y las mallas, en éstas últimas varias cadenas de aminoácidos se orientan y unen para formar una estructura plana. Todavía no se sabe si la estructura secundaria es consecuencia únicamente de la estructura primaria o si, en el momento de la síntesis proteica intervienen otros factores que contribuyen a que la proteína se doble de tal o cuál manera. Se piensa que el hecho de que la proteína sea ensamblada de un extremo a otro y no en bloques es un factor determinante para la estructura final de ésta.

El ADN tiene una estructura primaria parecida a la de las proteínas, pero tiene solamente cuatro tipos diferentes de grupos laterales (Timina,

Forma General

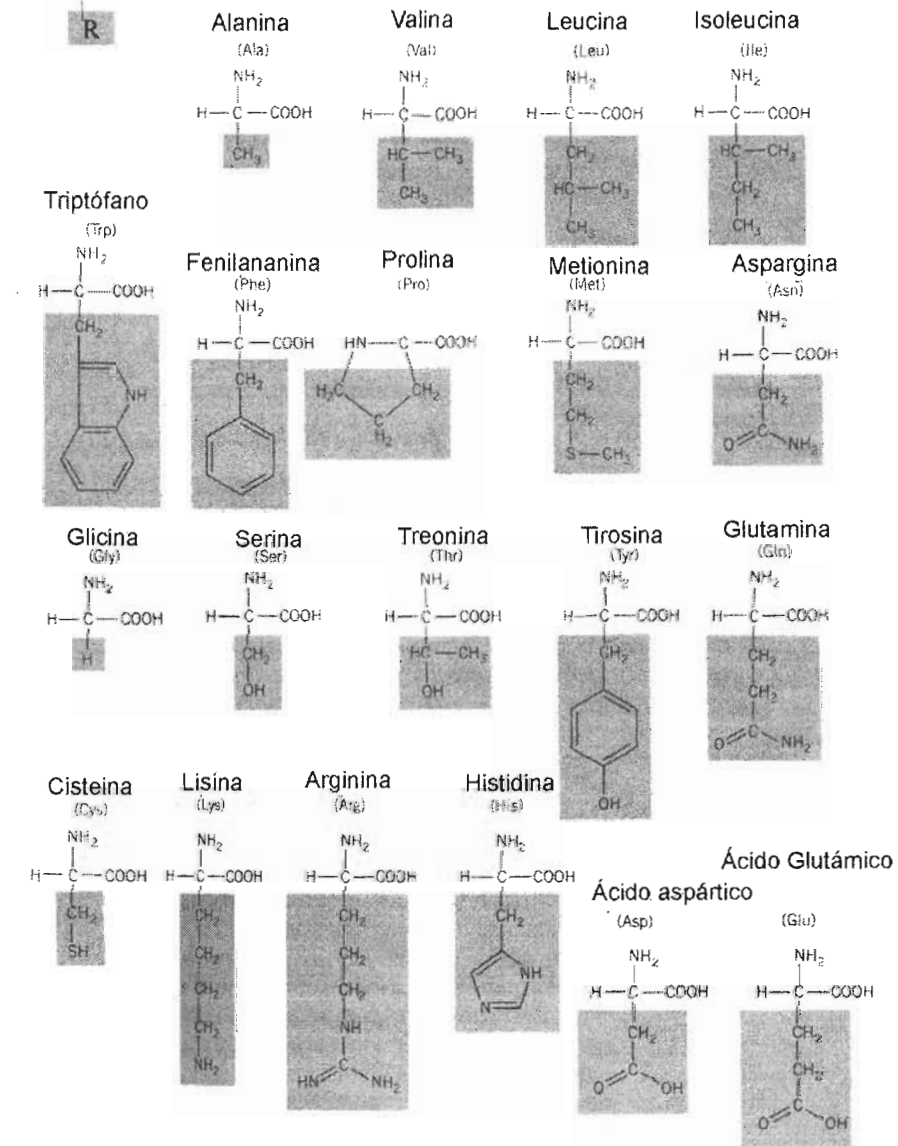
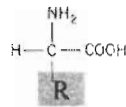


Figura 2.1: Aminoácido forma gal y grupos laterales

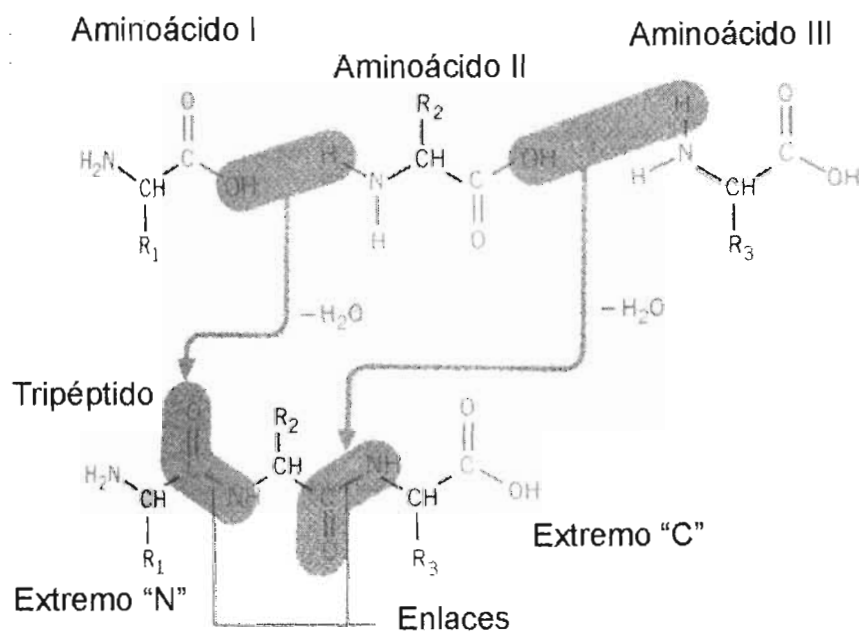


Figura 2.2: cadena de aminoácidos

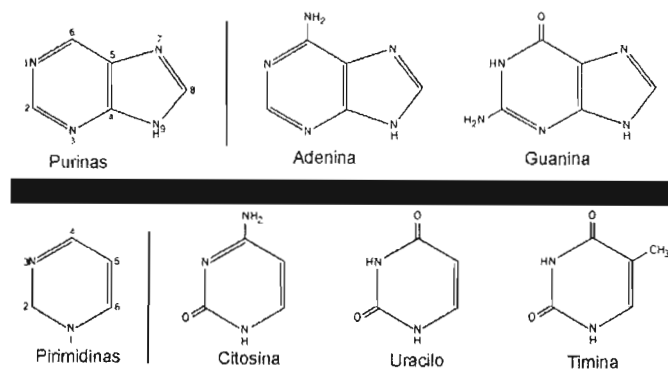


Figura 2.3: grupos laterales

Adenina, Guanina y Citosina) y se le conoce solamente un tipo de estructura secundaria: la doble hélice. En la doble hélice encontramos que existen dos cadenas apareadas, estas cadenas tienen por columna central a una repetición de desoxirribosas unidas por puentes de fosfato, a cada una de las desoxirribosas está unida una base (grupo lateral). Las cadenas se unen ligando sus grupos laterales, de una manera específica tal, que una Adenina en una cadena siempre estará ligada a una Timina en la otra cadena, y las Guaninas siempre estarán apareadas con Citosinas. A diferencia de las proteínas, el ADN si está totalmente caracterizado por su estructura primaria, y gracias al apareamiento unívoco entre purinas (Adeninas y Guaninas) y pirimidinas (Timinas y Citosinas) solo es necesario conocer la estructura primaria de una de las dos cadenas de la hélice para determinar por completo la estructura de ambas cadenas (ver Fig 2.4).

El ADN codifica información elemental para el organismo en su estructura primaria, es decir, en su secuencia característica de bases nitrogenadas. Toda esta información se almacena en unidades discretas llamadas cromosomas. Cada cromosoma consta de una única cadena ininterrumpida de ADN, que se divide en regiones de dos tipos, ADN *codificador* y ADN *no-codificador*; la diferencia entre estos dos tipos de regiones radica en que las cadenas que codifican tienen una utilidad clara de almacenamiento de información, mientras que pareciera ser que toda la otra región (la que no codifica) tienen una utilidad secundaria, posiblemente mecánica, o inclusive podría tratarse de información espúrea acumulada a través del tiempo.

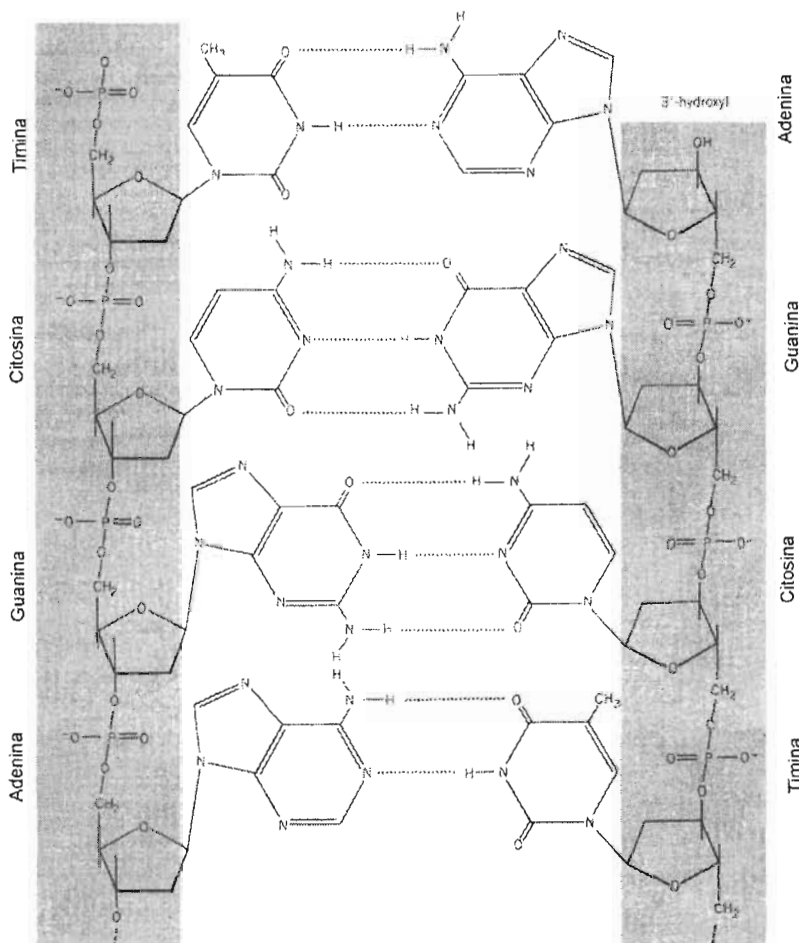


Figura 2.4: doble hélice

Los organismos sin núcleo (procariontes) tienen una proporción mucho menor de DNA no-codificador que los organismos con núcleo bien definido (eucariontes). Se pensaba que esta diferencia de proporciones se debía a que las células eucariontes habían acumulado una mayor cantidad de información espúrea dada una trayectoria mas larga de evolución, pero existen ahora opiniones diferentes que proponen que quizás las células procariontes también llegaron a tener una gran proporción de cadenas no-codificadoras, y que fué después de un proceso de depuración genética que lograron su estado actual.

Las cadenas codificadoras del ADN contienen en su secuencia primaria la información de la estructura de las proteínas que el organismo puede sintetizar, el código bajo el cuál realizan esta abstracción es bien conocido: El ADN tiene ,para cada una de sus cadenas, un *marco de lectura* bien definido (ver Fig 2.5), en donde cada terna de símbolos consecutivos son agrupados, para formar un código en base 64 (dado que son ternas de símbolos tomados de un conjunto de cuatro,  $4^3 = 64$ ), a cada una de éstas ternas de bases nitrogenadas se le llama *codón*. Se sabe que cada una de estas 64 posibles ternas o codones representan a uno de los 20 aminoácidos conocidos, por supuesto esta relación es una asignación degenerada, ya que cada codón representa a un solo aminoácido pero un aminoácido puede ser representado por varios codones diferentes.

Si entonces reemplazamos en la secuencia del ácido nucléico a los codones por los aminoácidos que representan, obtenemos una secuencia de aminoácidos, que resulta concordar con la secuencia de aminoácidos que forman a la proteína que es representada en el segmento de ADN del cuál se tomó la secuencia de codones.

## 2.2 TPBF y secuencias de codones

Cada organismo tiene una secuencia primaria diferente, inclusive dos organismos homocigóticos (engendrados a partir de una misma célula germinal), cuyo ejemplo por excelencia son los gemelos idénticos, tienen ciertas diferencias. Sin embargo las diferencias entre las secuencias genéticas de dos individuos de una misma especie son mínimas y se encuentran localizadas en regiones que constituyen, a lo más, un .1% de toda la cadena. Por otro lado, las diferencias entre dos especies diferentes pueden ser muy grandes. Para empezar, el ADN puede tener, dependiendo de la especie, extensiones (en secuencia) que van desde unos cuantos cientos de pares de bases, hasta

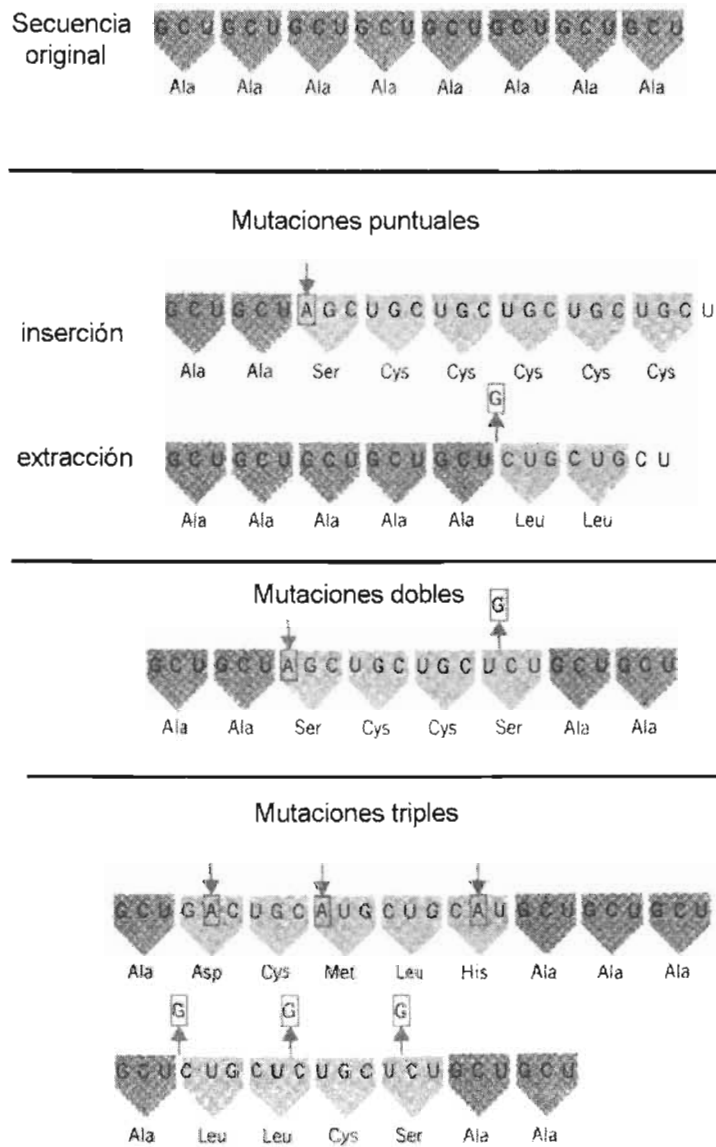


Figura 2.5: marco de lectura y codones



poco menos de  $10^{11}$  pares de bases en el caso de las plantas, pasando por un promedio de  $10^8$  pares de bases para los mamíferos. Los dos principales responsables de la evidente diversidad de secuencias diferentes son la mutación y la recombinación genética. El primero es un *accidente* que puede cambiar una secuencia genética de muchas maneras diferentes y es importante cuando dicho cambio trasciende al individuo y se perpetúa a través de las subsecuentes generaciones. El segundo, la recombinación genética, es un mecanismo natural en el cuál las características a nivel de genes de dos organismos progenitores son combinadas de diferentes maneras para lograr un tercer conjunto de características. La estructura básica de la secuencia de ADN total propia de una especie es entonces el resultado de una infinidad de cambios dados a través de millones de años, filtrados por los mecanismos de selección natural Darwiniana, y a pesar de que cada uno de estos cambios depende de una cantidad incalculable de factores, el resultado es de una naturaleza más sencilla de lo que se podría esperar.

Partiendo del hecho de que la distribución (en cantidad y posición) de cada una de las bases nitrogenadas que conforman las cadenas de información genética ha sido elegida de manera diferente para cada especie (y en menor grado para cada individuo) mediante un inmenso historial de cambios, cuya diversidad es tan grande como evidente es la de las especies que conocemos, es de suponer que, entre muchas otras cosas, la distribución rango-tamaño de las secuencias de codones varíe enormemente entre especies, cualitativa y cuantitativamente. Sorprendentemente no es así, existen dos formas generales posibles para la distribución rango-tamaño de codones, una propia de procariontes (ver Figs. 2.6 y 2.7) y otra para eucariontes.

En organismos procariontes el espectro de frecuencias se ajusta muy bien a una distribución Beta de dos parámetros (ec.1.18), una vez que normalizamos el dominio del histograma usando la expresión:

$$f = x^\alpha(N + 1 - x)^\beta \quad (2.1)$$

en donde  $N$  es el número de n-adas (codones) diferentes. En organismos eucariontes se observa una distribución más complicada (ver fig 2.7), esto se debe a que las cadenas de ADN de los organismos eucariontes están formadas por secuencias codificadoras (exones) alternadas de secuencias no codificadoras (intrones), ambas de naturalezas muy diferentes. La distribución rango tamaño de una muestra completa de eucarionte como la mostrada en la figura 2.7 se puede dividir en dos regiones, cada región se ajusta bien a una **TPBF**.

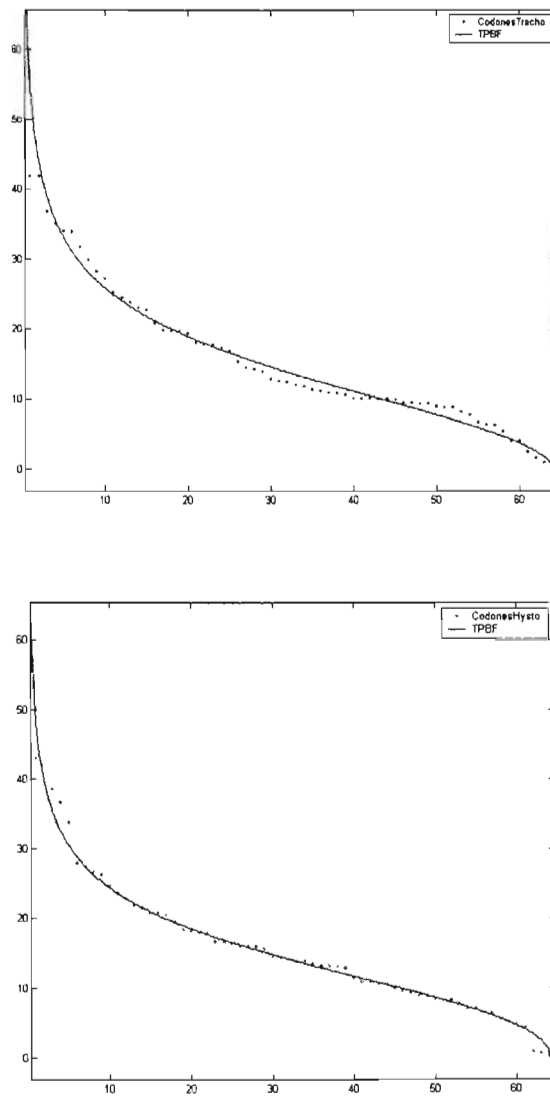


Figura 2.6: Distribución de codones en procariontes. Ejemplos: E.Hysto ( $\alpha = -.304, \beta = .351$ ). y CH. Tracho ( $\alpha = -.143, \beta = .813$ ).

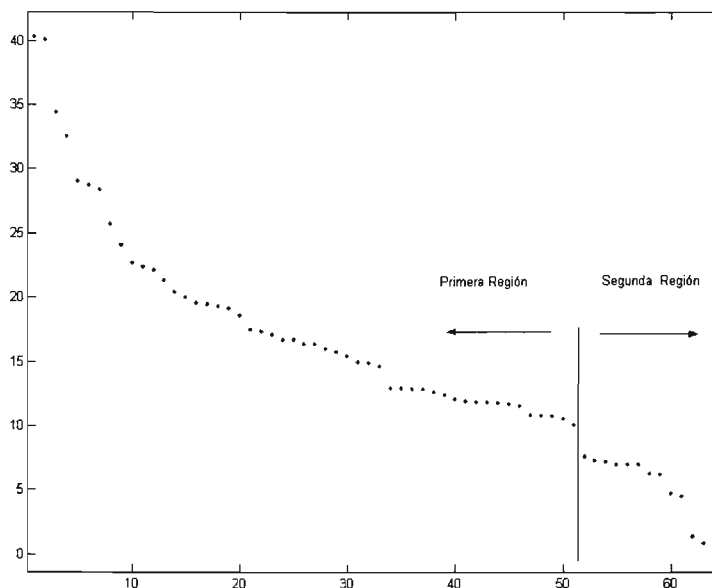


Figura 2.7: Distribución de codones en eucariontes. Ejemplo: *H. Sapiens*

La existencia de dos regiones se puede explicar si se toma en cuenta que una gran proporción (11% en *homo sapiens*) de la longitud de la secuencia total de ADN se compone de millones de repeticiones de sub-unidades pequeñas ( $\approx 300pb$ ) llamadas *SINE* (*short interspersed elements*). Los *SINE* son *retrotransposones*, secuencias que tienen la capacidad de copiarse en cadenas de mRNA y utilizar retrotranscriptasa para insertarse después en algún otro punto de la cadena. Este mecanismo difiere totalmente de la mecánica habitual de mutación de las secuencias codificadoras, y si existe un conjunto pequeño de codones usados preferentemente en estas secuencias, entonces la segunda región de la distribución reflejaría dicha preferencia.

## 2.3 TPBF y n-adas en secuencias de ADN

Si desplazamos el marco de lectura una o dos bases, y hacemos el mismo conteo para la distribución rango-tamaño encontramos de nuevo la distribución

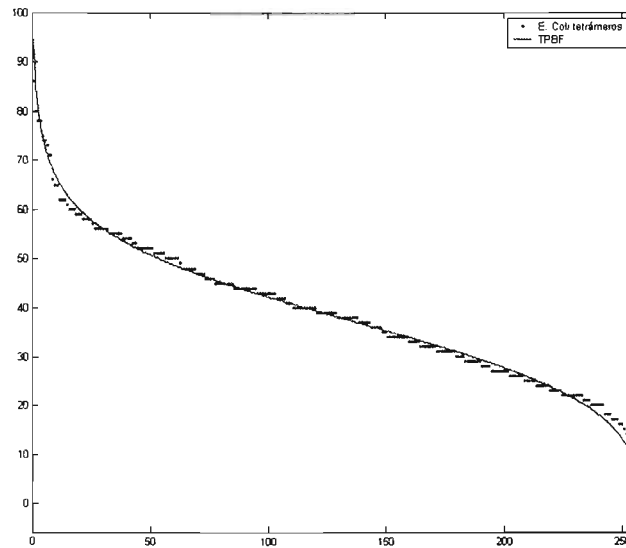


Figura 2.8: Distribución de grupos de 4 bases en E. Coli, ( $\alpha = -.137, \beta = 1.024$ )

1.18, e inclusive, si consideramos grupos de más de tres bases (n-adas) también encontramos una distribución de función Beta de dos parámetros (Figs. 2.8 y 2.9). Cuando consideramos codones, la distribución tenía solamente 64 ( $4^3$ ) valores distintos, un estudio con n-adas provee un histograma con mejor resolución, por ejemplo en la figura 2.8.

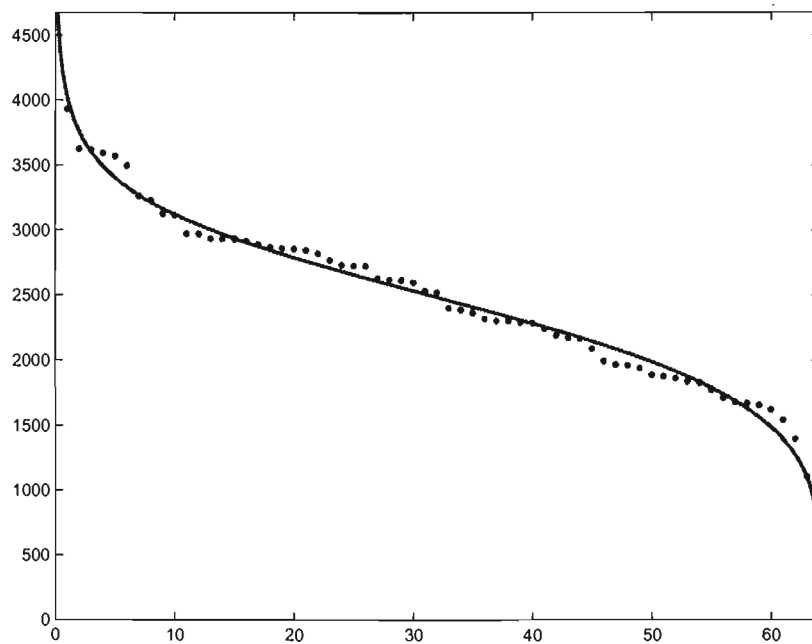


Figura 2.9: Distribución para grupos de tres bases con el marco de lectura desplazado. La curva continua es el ajuste con TPBF, el valor RMS para el ajuste es  $\approx .95999$

# Capítulo 3

## Recursos de la Dinámica Simbólica

### 3.1 Secuencias Simbólicas

Dado que las secuencias genéticas se manejan como secuencias simbólicas (de un lenguaje de cuatro letras: A, G, T, C), resulta natural referirnos a la dinámica simbólica para encontrar sistemas con propiedades similares a las estudiadas en los capítulos anteriores. Los mapeos discretos resultan ser una buena manera de generar secuencias simbólicas, y éstas han sido ampliamente estudiadas con relación a las propiedades caóticas de los mapeos que las producen. La manera más común de generar una secuencia simbólica es dividir el dominio continuo del mapeo en regiones discretas, asignándole a cada una de ellas un símbolo, de manera que una secuencia de iteraciones consecutivas se interpreta como una secuencia de símbolos. El ejemplo más famoso es probablemente el de la serie "U" o serie Universal, que descubrieron Metropolis et al. ([14]), para la clase de mapeos  $f(x)$  que cumplen:

$f(x)$  es derivable por trozos

$f(x) : [0, 1] \longrightarrow [0, 1]$  de manera continua

$$f(0) = f(1) = 0$$

$\exists! x_{\max} \in [0, 1]$  tal que  $f(x_{\max}) > f(y) \quad \forall y \in [0, 1], y \neq x_{\max}$

$f|_{x \leq x_{\max}}$  es monótonamente creciente

$f|_{x \geq x_{\max}}$  es monótonamente decreciente

$f(x)$  es derivable en  $x_{\max}$

Clase que incluye al mapeo logístico. La serie "U" se construye dividiendo el intervalo  $[0, 1]$  en dos regiones,  $[0, x_{\max}]$  y  $[x_{\max}, 1]$ , a la primera se le asigna el símbolo "L" y a la otra el símbolo "R". Ya hecha la convención se itera el valor  $x_{\max}$  para construir una secuencia de "R"s y "L"s. Lo que Metropolis *et al.* descubrieron fue que si se hace una lista ordenando las secuencias periódicas con respecto al parámetro del mapeo, ésta resulta ser la misma para todos los mapeos de la clase antes descrita. Encontraron también una manera de construir todas las secuencias periódicas de cualquier orden que se encuentran entre dos secuencias correspondientes a distintos valores del parámetro.

Si consideramos la curva logística (ec. 1.9) y hacemos la partición del intervalo antes descrita  $[0, \frac{1}{2}] - [\frac{1}{2}, 1]$  podemos construir series binarias como las de la secuencia universal. Si ahora agrupamos los símbolos subsecuentes en grupos de  $n$  tenemos una nueva serie cuyos elementos se toman de un grupo de  $2^n$ . Ahora podemos ordenar por rango a éstos  $2^n$  diferentes símbolos para obtener la distribución rango-tamaño propia de la secuencia.

### 3.2 n-ómeros de la curva logística y TPBF

Dependiendo del valor del parámetro  $\mu$  las secuencias de grupos de  $n$  pueden ser, desde una secuencia de un solo símbolo, en el caso de valores de  $\mu$  en donde el mapeo no se comporte de manera caótica, hasta una combinación caótica de los  $2^n$  símbolos diferentes, pasando por los patrones de la secuencia "U", si el valor de  $\mu$  corresponde a alguna ventana periódica. Si calculamos la distribución rango-tamaño de secuencias con un valor de  $\mu$  en la región no caótica obtendremos datos solo para unos cuantos símbolos, pero si acercamos el valor de  $\mu$  al valor de ergodicidad total ( $\mu \approx 4$ ), entonces en la distribución rango-tamaño empiezan a aparecer todos los símbolos posibles, y siempre tienen la misma forma funcional para la distribución: la función Beta de dos parámetros.

Los cálculos numéricos se hicieron en secuencias binarias ("0" para  $x \in [0, \frac{1}{2}]$  y "1" para  $x \in [\frac{1}{2}, 1]$ ) de 128,000 iteraciones del mapeo logístico con diferentes valores de  $\mu$  para obtener una serie de 16,000 grupos de 8 cifras binarias, cada uno correspondiendo a un entero entre 255 y 0. Se calculó el

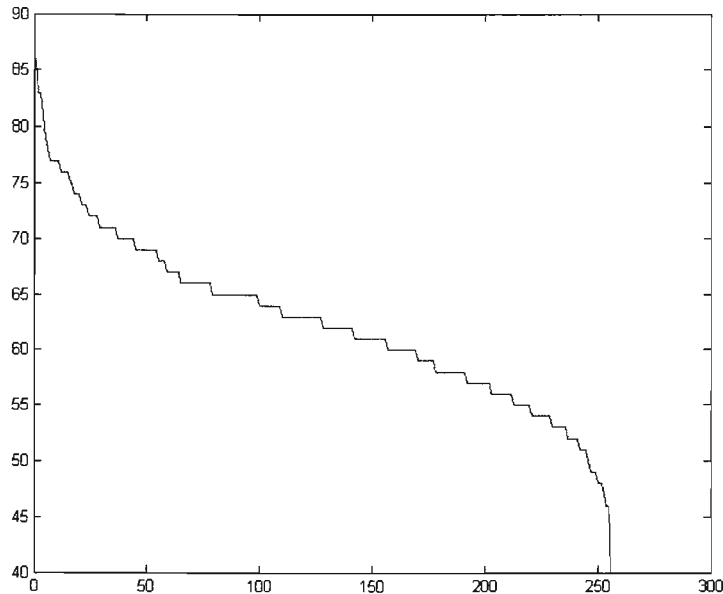


Figura 3.1: Distribución rango tamaño de 128,000 iteraciones en grupos de cuatro del mapeo logístico, para  $\mu = 4$  y  $\lambda = 0.6925$

exponente de Lyapunov ( $\lambda$ ) correspondiente a cada valor de  $\mu$ , para caracterizar el grado de caos en el sistema.

Para ajustar las curvas se eliminaron todos los valores iguales a cero exceptuando el primero, a manera de tener una distribución Beta completa, sin una "cola" de ceros en los valores bajos del rango. La ecuación para el ajuste no es precisamente la distribución Beta normalizada:

$$\rho(x) = \frac{x^\alpha(1-x)^\beta}{B(\alpha+1, \beta+1)} \quad (3.1)$$

dado que nuestros valores abarcan un dominio mayor que el intervalo  $[0, 1]$ . En vez de eso utilizamos la misma forma funcional pero renormalizada para ajustar al dominio de los datos:

$$\rho(x) = A[x^\alpha(N-x)^\beta] \quad (3.2)$$

en donde  $N-1$  es el número de valores del rango cuya frecuencia es diferente



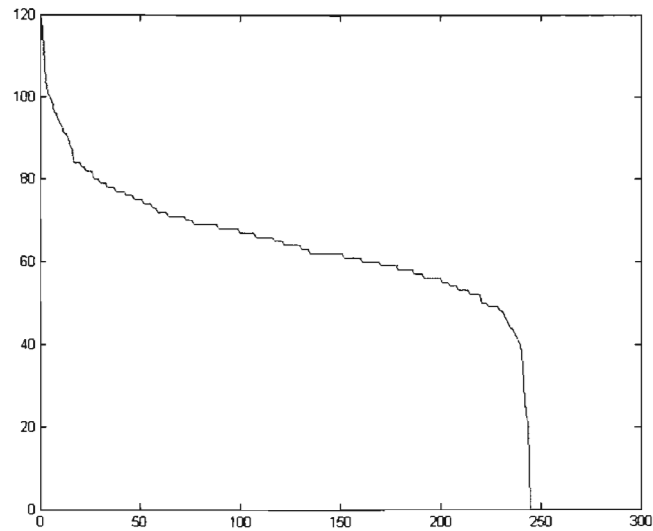


Figura 3.2: Distribución rango tamaño de 128,000 iteraciones en grupos de cuatro del mapeo logístico, para  $\mu = 3.99794661$  y  $\lambda = 0.669156521$

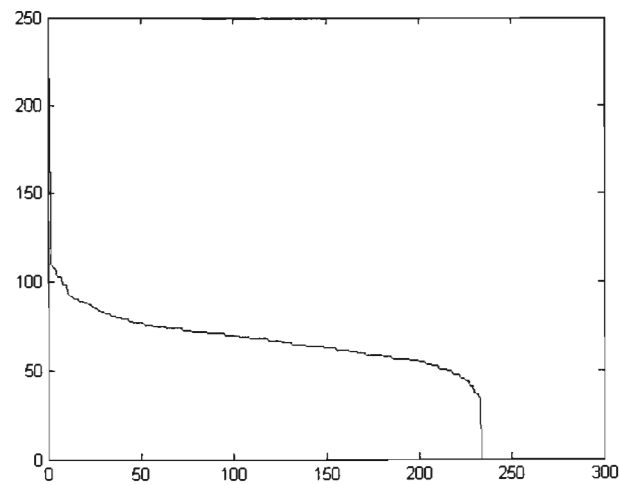


Figura 3.3: Distribución rango tamaño de 128,000 iteraciones en grupos de cuatro del mapeo logístico, para  $\mu = 3.99589322$  y  $\lambda = 0.662762179$

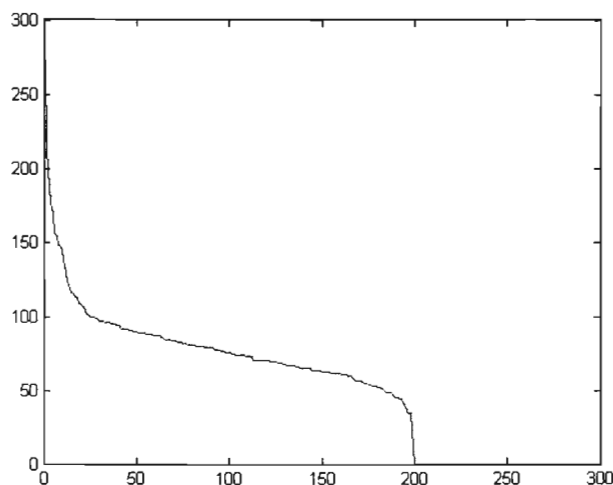


Figura 3.4: Distribución rango tamaño de 128,000 iteraciones en grupos de cuatro del mapeo logístico, para  $\mu = 3.98151951$  y  $\lambda = 0.618069$

de cero. El valor RMS de bondad para cada ajuste se muestra en la gráfica correspondiente (Figs. 3.5 a 3.7).

### 3.3 Modelo de expansión modificación

El modelo de expansión modificación (MEM) es un sistema dinámico simbólico que genera secuencias binarias mediante mecanismos que imitan las maneras más comunes de mutaciones en una secuencia genética. La manera de generar secuencias del modelo es la siguiente:

1. Se da una semilla arbitraria-De manera arbitraria se de una secuencia binaria pequeña, de dos o tres elementos.
2. Empieza la iteración-Cada elemento de la secuencia es expuesto al siguiente proceso: con una probabilidad  $p$  el elemento cambia al valor opuesto ( $1 \rightarrow 0$  ó  $0 \rightarrow 1$ ), en caso de que no se cumpla el evento de probabilidad  $p$  (con una probabilidad  $1 - p$ ) el elemento no cambia, se añade un elemento idéntico en el siguiente espacio de la secuencia

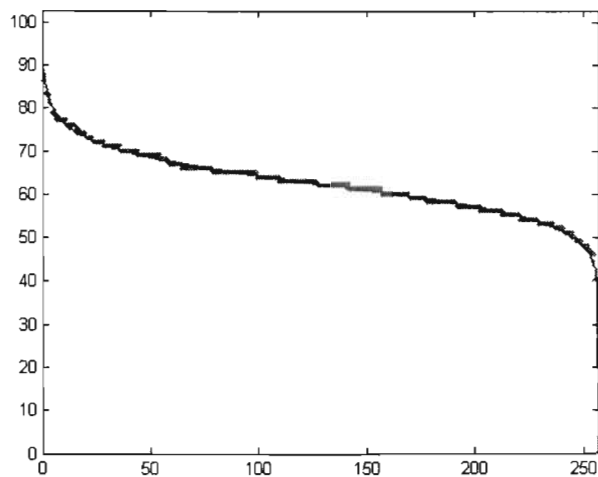


Figura 3.5: Ajuste para la muestra con  $\mu = 4$  y  $\lambda = 0.6925$ , el valor de bondad RMS=0.99615

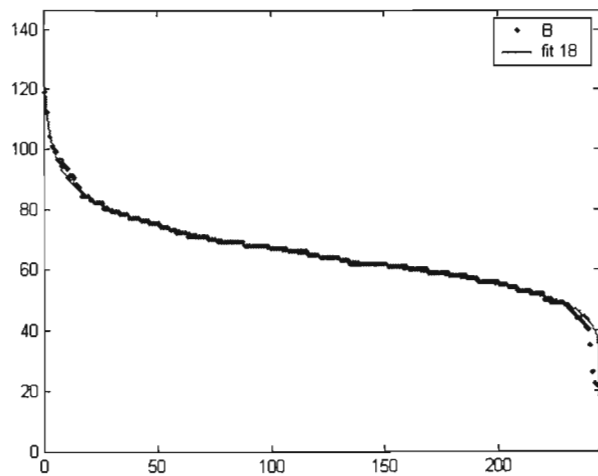


Figura 3.6: Ajuste para la muestra con  $\mu = 3.99794661$  y  $\lambda = 0.6925$ , el valor de bondad RMS=0.99787

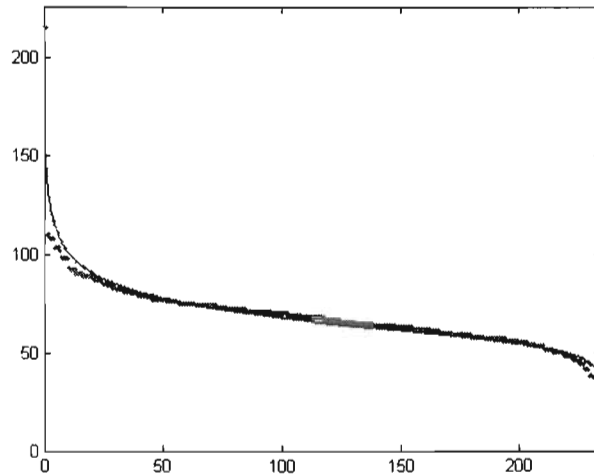


Figura 3.7: Ajuste para la muestra con  $\mu = 3.99589322$  y  $\lambda = 0.669156521$ , el valor de bondad  $RMS=0.90104$

3. Se agranda la secuencia-Repitiendo el procedimiento se agranda la secuencia, para lograr una distribución rango-tamaño de grupos de  $n$  elementos consecutivos es suficiente una longitud de  $n1000$ .

Por ejemplo, escogemos que la probabilidad de cambio sea  $p = .6$  empezamos con la semilla:

1101

para cada elemento se escoge un número aleatorio  $\epsilon \in [0, 1]$ .

.4753 .9541 .2168 .8743  
 $\underbrace{\hspace{1em}}_1 \quad \underbrace{\hspace{1em}}_1 \quad \underbrace{\hspace{1em}}_0 \quad \underbrace{\hspace{1em}}_1$

Si  $\epsilon$  para un elemento resulta ser  $\epsilon < .6$  entonces cambiamos el valor del elemento, de otra manera se inserta un elemento igual justo en el espacio siguiente:

.4753 .9541 .2168 .8743  
 $\underbrace{\hspace{1em}}_1 \quad \underbrace{\hspace{1em}}_1 \quad \underbrace{\hspace{1em}}_0 \quad \underbrace{\hspace{1em}}_1$   
 $\rightarrow$   
 $\underbrace{\hspace{1em}}_0 \quad \underbrace{\hspace{1em}}_{11} \quad \underbrace{\hspace{1em}}_1 \quad \underbrace{\hspace{1em}}_{11}$

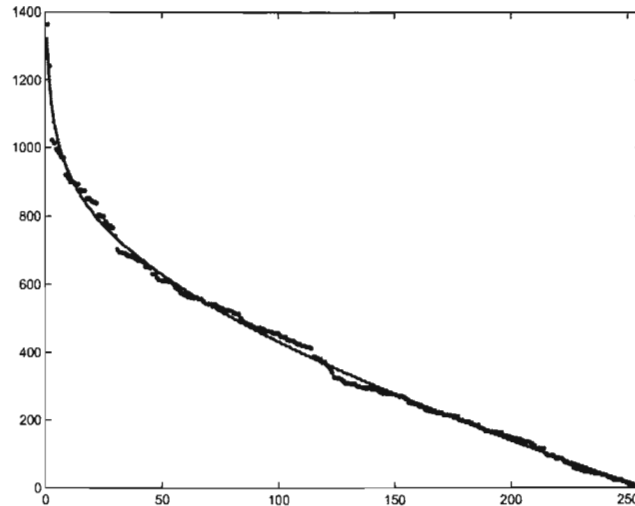


Figura 3.8: Distribución rango-tamaño de una secuencia generada por modelo de expansión-modificación para una probabilidad de cambio  $p = .9$  y grupos de ocho elementos. El valor RMS para el ajuste (línea continua en la gráfica) es de .996.

Para entonces acabar con la secuencia:

011111

y volver a iterar hasta lograr una secuencia grande.

La distribución rango-tamaño de una de estas secuencias, sin importar del valor del parámetro  $p$  o el número  $n$  para el tamaño de los grupos a contar, es ajustada con éxito <sup>1</sup> por una distribución Beta de dos parámetros.

---

<sup>1</sup>El valor RMS para los ajustes rara vez difería de 1 por más de una centésima.

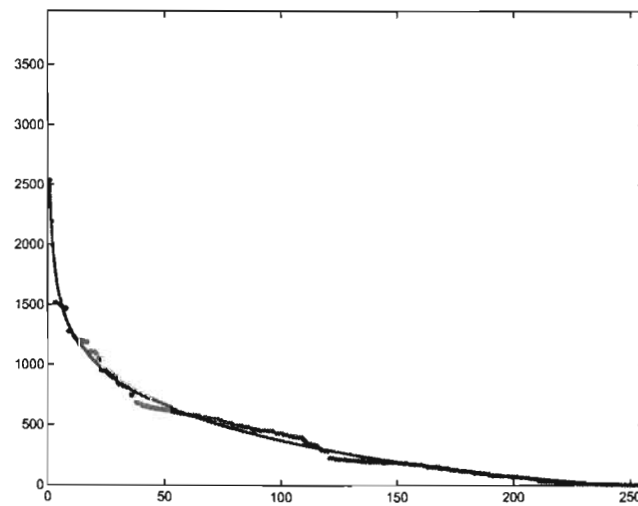


Figura 3.9: Distribución rango-tamaño de una secuencia generada por modelo de expansión-modificación para una probabilidad de cambio  $p = .5$  y grupos de ocho elementos. El valor RMS para el ajuste (línea continua en la gráfica) es de .98896.

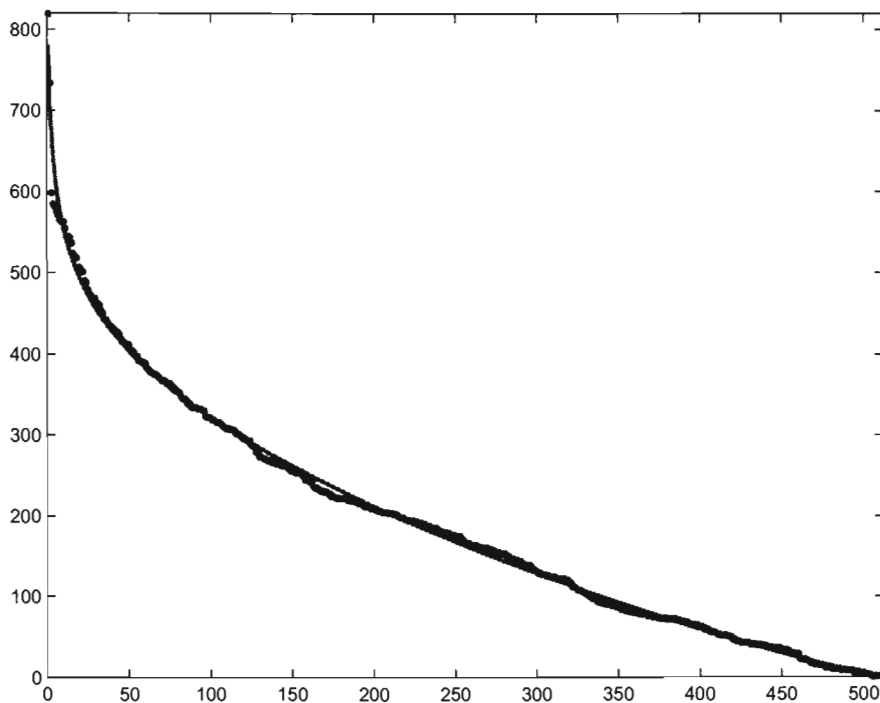


Figura 3.10: Distribución rango-tamaño de una secuencia generada por modelo de expansión-modificación para una probabilidad de cambio  $p = .9$  y grupos de nueve elementos. El valor RMS para el ajuste (línea continua en la gráfica) es de .99756.

# Capítulo 4

## Inversión de Frobenius Perron y TPBF.

### 4.1 Método de Conjugación

El estudio de las propiedades estadísticas de un sistema puede llevar a primeras ideas acerca de la naturaleza íntima del sistema. En el primer capítulo vimos como a cada mapeo le podemos asignar una densidad de probabilidad, utilizando la ecuación de Frobenius-Perron (1.8), un proceso inverso puede construir un mapeo caótico a partir de propiedades estadísticas arbitrarias, tales como la densidad de probabilidad o la función de correlación. A dicho proceso se le llama problema inverso de Frobenius-Perron. Muchos métodos diversos para generar mapeos que cumplan propiedades estadísticas dadas se han publicado ([7, 8, 10, 13]), uno que nos es particularmente útil es el método de conjugación (*Conjugation Transformation*).

Dado que la elección de densidad de probabilidad no determina unívocamente a un mapeo, cada método exige la elección arbitraria de otras características para crear al mapeo en cuestión. El método por conjugación consiste en la composición de funciones de distribución acumulada determinada con un mapeo de densidad de probabilidad constante arbitrario, las funciones de distribución imponen al mapeo la densidad de probabilidad que deseamos, mientras que la elección arbitraria del mapeo con densidad uniforme determina su forma funcional. Consideremos un par de mapeos:

$$\psi : \mathbf{Y} \longrightarrow \mathbf{Y} \tag{4.1}$$



y:

$$\phi : \mathbf{Z} \longrightarrow \mathbf{Z} \tag{4.2}$$

Tales que exista un homeomorfismo  $h : \mathbf{Z} \longrightarrow \mathbf{Y}$  que cumpla:

$$h \circ \phi = \psi \circ h \tag{4.3}$$

Es decir  $\psi$  y  $\phi$  son mapeos homeomorfos. Si pedimos que  $y_0 = h(z_0)$  entonces:

$$y_n = h(z_n). \quad \forall n \tag{4.4}$$

Si además suponemos que  $\mathbf{Z} = [0, 1]$  y que  $\phi$  tiene una función de probabilidad homogénea, es decir, que los iterados  $z_0, z_1, \dots, z_n \dots$  están distribuidos uniformemente en  $\mathbf{Z}$  (En el sentido de grandes números, por supuesto), tenemos que la variable aleatoria:

$$y_n = h(z_n) = h(\phi^n(z_0)) \tag{4.5}$$

tiene una distribución de probabilidad acumulada  $\mathbf{F} = h^{-1}$ , y entonces si queremos construir un mapeo  $\psi$  con una densidad de probabilidad  $\rho(x)$ , hacemos:

$$\mathbf{F}(x) = \int_0^x \rho(\xi) d\xi \tag{4.6}$$

y:

$$\psi = \mathbf{F}^{-1} \circ \phi \circ \mathbf{F} \tag{4.7}$$

Donde  $\phi$  es cualquier mapeo con una distribución homogénea, como por ejemplo, el mapeo 1.14.

Existe una manera alternativa de construir el mapeo con la densidad de probabilidad requerida ([10]), ahora en vez de escoger un mapeo arbitrario con densidad uniforme como en el método antes descrito, podemos imponer condiciones acerca de la simetría que le queremos dar al mapeo a construir. Empezamos con la forma diferencial de la ecuación de Frobenius Perron:

$$\rho(y)|dy| = \sum_{x_i=f^{-1}(y)} \rho(x_i)|dx_i| \tag{4.8}$$

donde los índices de la suma,  $x_i$ , son todas las preimágenes de  $y$ . Aquí consideramos mapeos unimodales  $f(x)$ , que obedecen las siguientes restricciones:

$$f(x) : [0, 1] \longrightarrow [0, 1] \quad \text{de manera continua}$$

$$f(0) = f(1) = 0$$

$$\exists! x_{\max} \in [0, 1] \text{ tal que } f(x_{\max}) > f(y) \quad \forall y \in [0, 1], y \neq x_{\max}$$

$$f|_{x \leq x_{\max}} \quad \text{es monótonamente creciente}$$

$$f|_{x \geq x_{\max}} \quad \text{es monótonamente decreciente}$$

Un mapeo de este estilo tiene siempre solamente dos preimágenes por cada valor, es decir  $\forall y \in [0, 1], \exists x_I, x_D$  con  $x_D \neq x_I$  tales que  $f(x_I) = y = f(x_D)$  y  $\forall z \neq x_I, \neq x_D \Rightarrow f(z) \neq y$ . Dado que  $f$  es continua existe una función  $h(x)$  que cumple lo siguiente:

$$h(x) : [0, x_{\max}] \rightarrow [x_{\max}, 1] \quad \text{de manera continua}$$

$$h(x_I) = x_D$$

Dado que solo existen dos preimágenes, la ec 4.8 tiene solo dos sumandos del lado derecho:

$$\rho(y)|dy| = \rho(x_I)|dx_I| + \rho(x_D)|dx_D| \quad (4.9)$$

Y sustituyendo la función  $h(x)$  obtenemos:

$$\rho(y)|dy| = \rho(x_I)|dx_I| + \rho(h(x_I))|h'(x_I)dx_I| \quad (4.10)$$

Si consideramos solo el lado izquierdo del mapeo ( $[0, x_{\max}]$ ) y tomamos en cuenta también que  $h'(x) < 0$ , obtenemos:

$$\rho(y)dy = \rho(x_I)dx_I - \rho(h(x_I))h'(x_I)dx_I \quad (4.11)$$

Hemos eliminado una diferencial de la ecuación y la integración es directa utilizando la definición de función de probabilidad acumulada:

$$\mu(x) = \int_0^x \rho(t)dt \quad (4.12)$$

La integral se hace en la sección izquierda tomando  $x < x_{\max}$ , donde la función  $h(x)$  está definida:

$$\int_0^{f(x)} \rho(t)dt = \int_0^x \rho(t)dt - \int_{h(0)}^{h(x)} \rho(t)dt \quad (4.13)$$

$$\mu(f(x)) = \mu(x) - \mu(h(x)) + \mu(h(0)) \quad (4.14)$$

Obviamente tenemos:

$$\mu(0) = 0 \quad (4.15)$$

y:

$$\mu(h(0)) = \mu(1) = 1 \quad (4.16)$$

La ecuación implícita queda como:

$$\mu(f(x)) = \mu(x) - \mu(h(x)) + 1 \quad (4.17)$$

Y finalmente resolviendo para  $f(x)$  (sin olvidar que la ecuación solo es válida en la región creciente) obtenemos:

$$f(x)|_{x < x_{\max}} = \mu^{-1}[\mu(x) - \mu(h(x)) + 1] \quad (4.18)$$

Para encontrar la forma funcional de la sección derecha ( $x \geq x_{\max}$ ) partimos de la ecuación 4.9, pero ahora sustituimos con:

$$x_I = h^{-1}(x_D) \quad (4.19)$$

en vez de:

$$h(x_I) = x_D$$

Y los análogos de las ecuaciones 4.9-4.18 son:

$$-\rho(y)dy = -\rho(h^{-1}(x_D))(h^{-1}(x_D))'dx_D + \rho(x_D)dx_D \quad (4.20)$$

$$-\int_{f(x)}^{f(1)} \rho(t)dt = -\int_{h^{-1}(x)}^{h^{-1}(1)} \rho(t)dt + \int_x^1 \rho(t)dt \quad (4.21)$$

$$-\mu(f(1)) + \mu(f(x)) = -\mu(h^{-1}(1)) + \mu(h^{-1}(x)) + \mu(1) - \mu(x) \quad (4.22)$$

$$-\mu(0) + \mu(f(x)) = -\mu(0) + \mu(h^{-1}(x)) + 1 - \mu(x) \quad (4.23)$$

$$\mu(f(x)) = \mu(h^{-1}(x)) - \mu(x) + 1 \quad (4.24)$$

Y finalmente:

$$f(x)|_{x \geq x_{\max}} = \mu^{-1}[-\mu(x) + \mu(h^{-1}(x)) + 1] \quad (4.25)$$

$$f(x)|_{x < x_{\max}} = \mu^{-1}[\mu(x) - \mu(h(x)) + 1] \quad (4.26)$$

En este método, de todo el conjunto de soluciones, se elige una al seleccionar las propiedades de simetría mediante la función  $h(x)$ . Si en particular decidimos que nuestro mapeo sea simétrico (con respecto al valor medio  $\frac{1}{2}$ ),

entonces hacemos que  $h(x) = 1 - x$ , y las formas funcionales 4.18 y 4.25 quedan:

$$f(x)|_{x \geq x_{\max}} = \mu^{-1}[-\mu(x) + \mu(1 - x) + 1] \quad (4.27)$$

$$f(x)|_{x < x_{\max}} = \mu^{-1}[\mu(x) - \mu(1 - x) + 1] \quad (4.28)$$

Y para unir estas dos formas notamos primero que:

$$x \geq x_{\max} = \frac{1}{2} \Rightarrow [-\mu(x) + \mu(1 - x)] < 0 \quad (4.29)$$

$$x < x_{\max} = \frac{1}{2} \Rightarrow [\mu(x) - \mu(1 - x)] < 0 \quad (4.30)$$

Puesto que  $\mu(x)$  es la integral definida de  $\rho(x)$  que cumple  $\rho(x) \geq 0$  en tanto que es una función de densidad de probabilidad. La solución a nuestro problema queda entonces como:

$$f(x) = \mu^{-1}[1 - |\mu(x) - \mu(1 - x)|] \quad (4.31)$$

Es posible también una solución asimétrica sencilla para reproducir la estadística de la función Beta de dos parámetros, pero se le exige al mapeo resultante que cumpla que:

$$\forall x_D \neq x_I, \text{ con: } f(x_D) = f(x_I), \Rightarrow \rho(x_D) = \rho(x_I) \quad (4.32)$$

Desafortunadamente esta última condición solo sucede si los dos parámetros de la función Beta,  $\alpha$  y  $\beta$  cumplen  $\alpha\beta > 0$ , mientras que los ajustes a los datos de codones que nos interesan todos resultan tener parámetros que siguen este patrón:

$$\begin{aligned} \alpha, \beta &< 1 \\ \alpha\beta &< 0 \end{aligned}$$

## 4.2 Problema inverso de FP aplicado a TPBF

Derivaremos, utilizando las técnicas de la sección pasada, mapeos que tengan una función Beta de dos parámetros por densidad de probabilidad, para luego identificar las propiedades del mapeo que provocan ésta densidad particular, con los mecanismos de los que se valen los tipos más comunes de mutaciones

para lograr el mismo patrón de distribución y con otros métodos de imitación estadística que serán mostrados más adelante.

En el primer método, en el caso asimétrico, encontramos que el mapeo se puede fabricar utilizando la ecuación 4.7:

$$\psi = \mathbf{F}^{-1} \circ \phi \circ \mathbf{F} \quad (4.33)$$

Donde  $F$  es la distribución de probabilidad acumulada:

$$F(x) = \int_0^x \rho(\xi) d\xi \quad (4.34)$$

En nuestro caso, la función de densidad de probabilidad es la función Beta de dos parámetros:

$$\rho(x) = \frac{x^\alpha(1-x)^\beta}{B(\alpha+1, \beta+1)} \quad (4.35)$$

con:

$$B(x, y) = \int_0^1 t^{x-1}(1-t)^{y-1} dt \quad (4.36)$$

y por lo tanto nuestra distribución de probabilidad acumulada es la función Beta incompleta normalizada:

$$B_N(x, a, b) = \frac{\int_0^x \xi^a(1-\xi)^b d\xi}{B(a+1, b+1)} \quad (4.37)$$

Cuidando la notación, identificamos los parámetros de la función Beta incompleta con los de la distribución Beta de dos parámetros:

$$\begin{aligned} a &\rightarrow \alpha + 1 \\ b &\rightarrow \beta + 1 \end{aligned}$$

Nuestra distribución queda entonces:

$$B_N(x, \alpha + 1, \beta + 1) = \frac{\int_0^x \xi^\alpha(1-\xi)^\beta d\xi}{B(\alpha + 1, \beta + 1)} \quad (4.38)$$

Ahora necesitamos un mapeo con densidad de probabilidad homogénea  $\phi(x)$ , utilizaremos el mapeo triangular:

$$\phi(x) = 2|x - \frac{1}{2}| + 1 \quad (4.39)$$

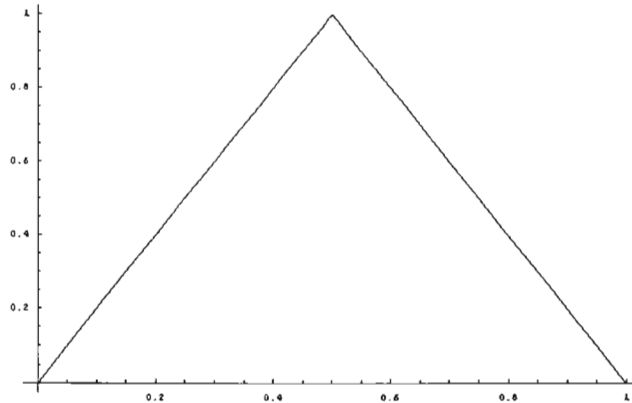


Figura 4.1: Mapeo triangular

La forma final del mapeo creado con el primer método queda:

$$\psi(x, \alpha, \beta) = B_N^{-1}(2|B_N(x, \alpha + 1, \beta + 1) - \frac{1}{2}| + 1, \alpha + 1, \beta + 1) \quad (4.40)$$

No tenemos una forma analítica para la ecuación anterior, pero afortunadamente existen muchos algoritmos numéricos para aproximar la función Beta.

El segundo método, el de solución simétrica tendría la forma:

$$\psi(x, \alpha, \beta) = B_N^{-1}(1 - |B_N(x, \alpha + 1, \beta + 1) - B_N(1 - x, \alpha + 1, \beta + 1)|, \alpha + 1, \beta + 1) \quad (4.41)$$

Conviene analizar el comportamiento cualitativo de la forma de los mapeos con respecto a sus dos parámetros, para poder identificarlos con los parámetros de los otros sistemas con distribución Beta que veamos mas adelante. Por ahora solo se muestran gráficas de dicha dependencia (Figs. 4.4, 4.5, 4.9 y 4.10).

En el capítulo 3, mediante el método inverso de Frobenius Perron, se fabricaron dos mapeos dependientes ambos de dos parámetros  $(\alpha, \beta)$ , que tienen por densidad de probabilidad la distribución Beta con los mismos dos parámetros. Para entender la naturaleza de la distribución Beta podemos prestarle atención a los cambios en la forma funcional de los mapeos cuando los parámetros de la función Beta son alterados, ya que en ellos se verán reflejados los mecanismos de distribución de los iterados, y por lo tanto, la forma cualitativa de la densidad de probabilidad.

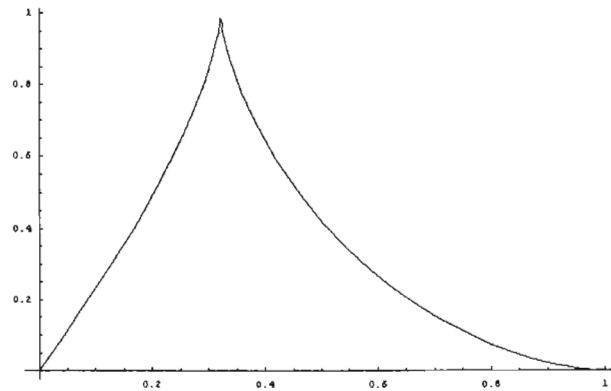


Figura 4.2: Solución asimétrica del problema inverso FP ( $\alpha = -.134, \beta = .513$ )

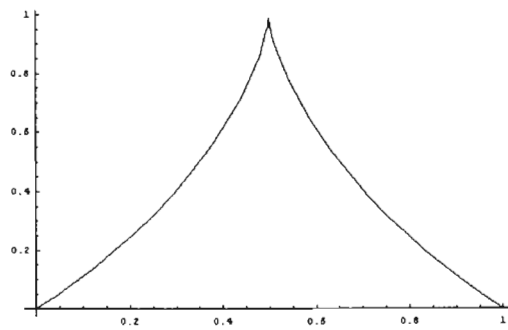


Figura 4.3: Solución simétrica del problema inverso FP ( $\alpha = -.134, \beta = .513$ )

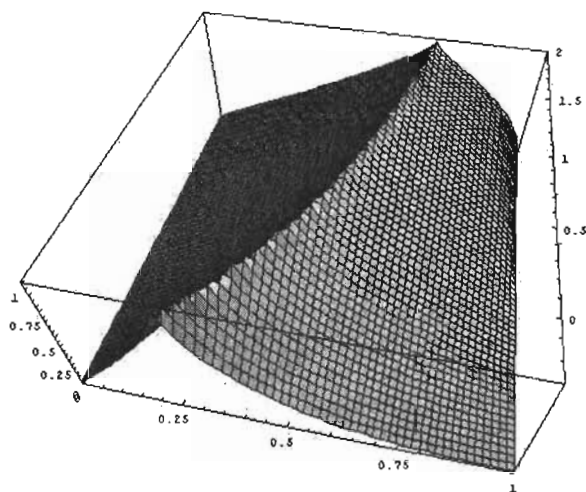


Figura 4.4: Dependencia del parámetro  $\alpha$  en el rango  $[-.3,2]$  para la solución asimétrica

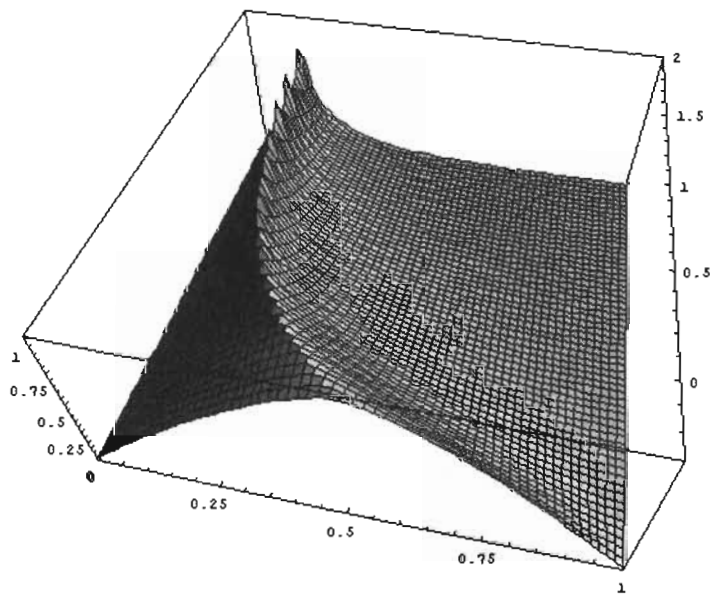


Figura 4.5: Dependencia del parámetro  $\beta$  en el rango  $[-.3,2]$  para la solución asimétrica



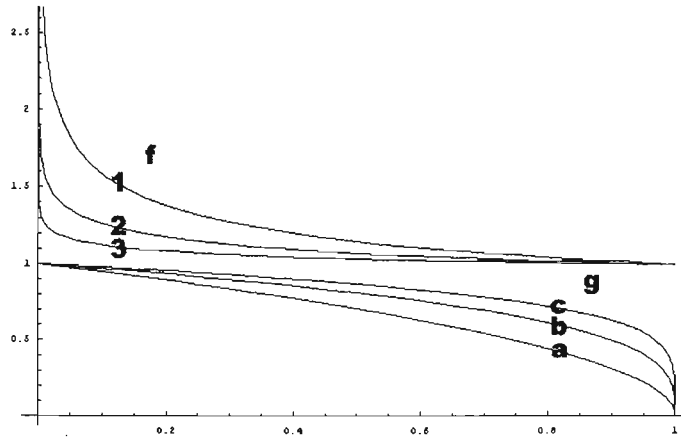


Figura 4.6: Factores de la distribución Beta, la función  $f$  con parámetro 1,  $\alpha = -0.1$ , 2,  $\alpha = -0.2$  y 3,  $\alpha = -0.05$  y la función  $g$  con parámetro  $a, \beta = 0.5$ ,  $b, \beta = 0.3$  y  $c, \beta = 0.2$

La distribución Beta es el producto de dos funciones sencillas,  $f = x^\alpha$  y  $g = (1 - x)^\beta$ , cada una depende de solo un parámetro. En la figura 4.6 podemos ver como la apariencia final del producto de las dos funciones está dominada en un extremo del intervalo por solo uno de los factores y en el otro extremo el dominio es del otro factor. El factor  $f$  domina sobre el extremo izquierdo, dado que en esa región el otro factor,  $g$ , tiende rápidamente a 1. De la misma manera la función  $g$  domina sobre el extremo derecho. El parámetro  $\beta$  tiene entonces influencia disrecta sobre la curvatura del extremo derecho de la distribución, mientras que la curvatura del lado izquierdo está determinada en gran medida por el parámetro  $\alpha$ , aunque por supuesto los dos parámetros tienen, aunque mínima, una influencia en ambas curvaturas.

Si aumentamos el valor del parámetro  $\beta$  en el rango  $[0, 1]$  vemos que en el extremo derecho la distribución aumenta su radio de curvatura para convertirse en un segmento que tiende hacia una línea recta con pendiente  $-1$ , si seguimos aumentando el valor del parámetro hasta el rango  $[1, \infty)$ , el segmento cambia de ser una curva cóncava a una convexa. El comportamiento en ambos casos es una disminución de la densidad de probabilidad en el extremo derecho, si la distribución está normalizada obtendremos un correspondiente aumento de la densidad de probabilidad en el extremo izquierdo.

Si ahora disminuimos el valor del parámetro  $\alpha$  en el rango  $[-1, 0]$  notamos

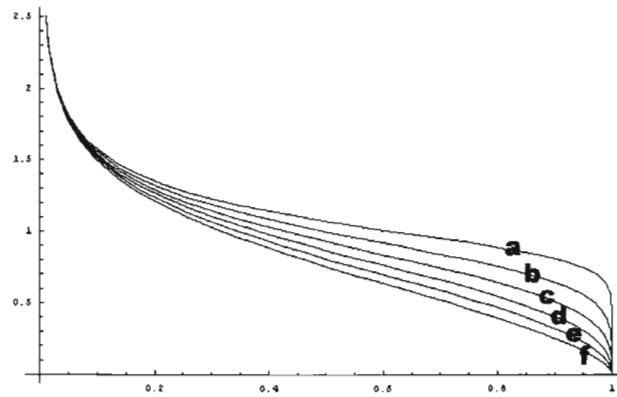


Figura 4.7: Dependencia de la función TPBF del parámetro  $\beta$ , los valores para cada curva son:  $a - \beta = .1, b - \beta = .2, c - \beta = .3, d - \beta = .3, e - \beta = .4, f - \beta = .5$  y para todas  $\alpha = -.2$

que el radio de curvatura disminuye, haciendo que la función converja de manera más lenta hacia infinito cuando  $x \rightarrow 0^-$ , aumentando el valor de la integral de probabilidad del lado izquierdo.

Los mapeos creados mediante método de FP sufrirán cambios ante un cambio en los parámetros de la distribución a partir de la cuál fueron creados, ahora que sabemos como depende la distribución Beta de sus parámetros podemos encontrar relación entre la forma de los mapeos y su correspondiente densidad.

En la solución simétrica un cambio en cualquiera de los dos parámetros tiene como efecto el alejamiento o acercamiento de la curva del mapeo a la recta identidad, cada parámetro lo hace de una manera ligeramente diferente, pero el resultado cualitativo es el mismo.

El resultado de que la función del mapeo se aproxime a una recta con pendiente uno es la de un desplazamiento hacia la izquierda de la densidad de probabilidad. Este es un hecho que se puede demostrar intuitivamente si consideramos el efecto que provoca la aproximación del mapeo a la identidad en los intervalos de la zona izquierda del dominio del mapeo que son iterados hacia la parte derecha del dominio. Una angostamiento entre la identidad y el mapeo (En todo momento la curva del mapeo permanece sobre la identidad) provoca un engrosamiento de los intervalos del lado izquierdo que son mapeados al intervalo izquierdo de nuevo, y el consecuente adelgazamiento de los intervalos del lado izquierdo que son mapeados fuera de éste. Este

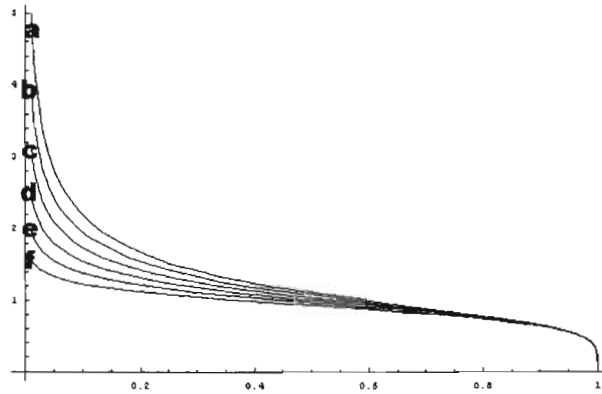


Figura 4.8: Dependencia de la función TPBF del parámetro  $\alpha$ , los valores para cada curva son:  $a - \alpha = -.35, b - \alpha = -.3, c - \alpha = -.25, d - \alpha = -.2, e - \alpha = -.15, f - \alpha = -.1$  y para todas  $\beta = -.2$

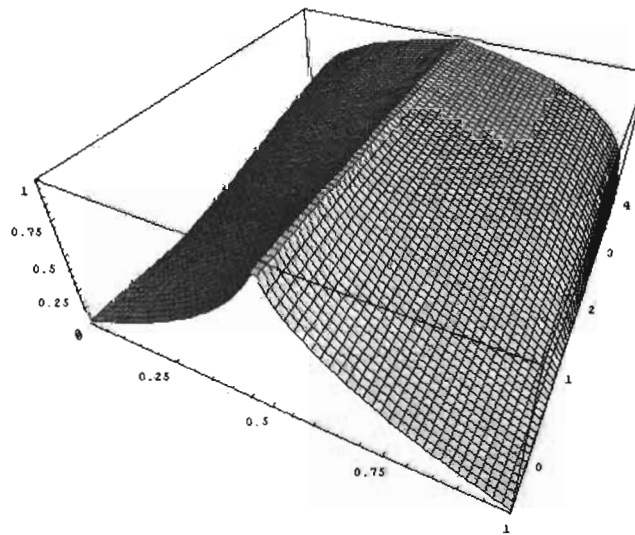


Figura 4.9: Dependencia del parámetro  $\alpha$  en el rango  $[-.5, 4]$  para la solución simétrica

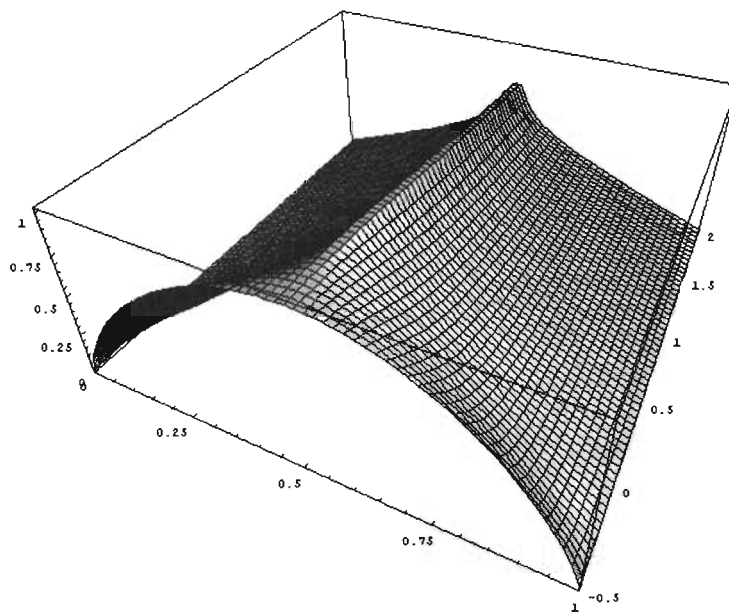


Figura 4.10: Dependencia del parámetro  $\beta$  en el rango  $[-.5,2]$  para la solución simétrica

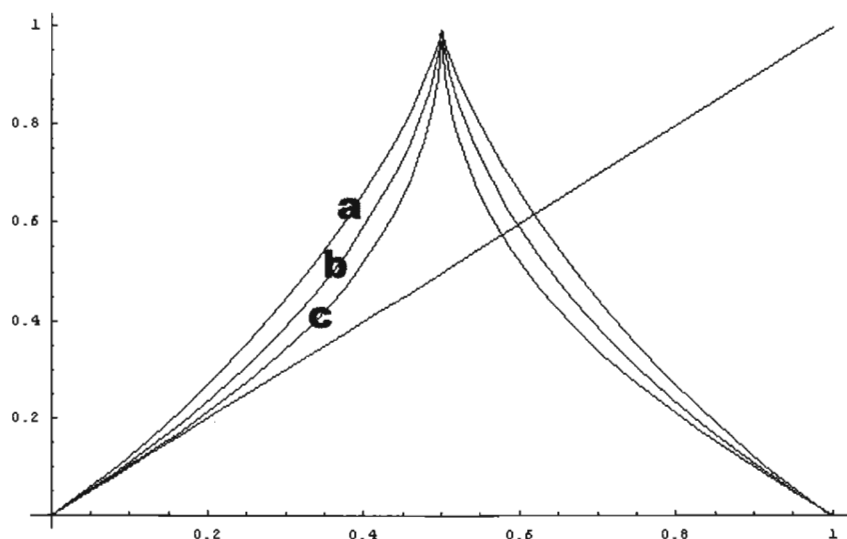


Figura 4.11: Angostamiento del mapeo cuando  $\beta \rightarrow 1^-$  y  $-\alpha \rightarrow 1^-$ . Los valores de los parámetros para las diferentes curvas son  $a \rightarrow [-\alpha = \beta = .3]$ ,  $b \rightarrow [-\alpha = \beta = .5]$ ,  $c \rightarrow [-\alpha = \beta = .8]$ .

engrosamiento se da también en los intervalos que permanecen en la mitad izquierda después de más de una iteración.

Cada parámetro cambia la forma del mapeo de una forma cualitativamente parecida, pero existe una sutileza que hace que los cambios que aparecen en el mapeo dado el cambio de uno u otro parámetro tengan repercusiones en extremos distintos de la distribución de probabilidad.

Los límites  $\alpha \rightarrow -1^+$  y  $\beta \rightarrow 1^-$  tienen por consecuencia el acercamiento de la curva del mapeo a la recta identidad y un angostamiento de la curva del mapeo en la vecindad del máximo, ver figs 4.12 y 4.13. La diferencia entre el cambio producido por una u otra variable reside en que el cambio en  $\alpha$  hace que el mapeo se aproxime a la identidad en el primer cuarto del intervalo, mientras que el cambio en  $\beta$  tiene un efecto notorio de acercar la curva del mapeo al eje de simetría vertical del mapeo.

La consecuencia de acercar la curva del mapeo al eje de simetría vertical es la disminución de la longitud diferencial del intervalo que es mapeado a un intervalo de longitud diferencial en el extremo derecho del dominio de ahí que el efecto del cambio del parámetro  $\beta$  en la distribución de probabilidad sea más notorio en la región derecha.

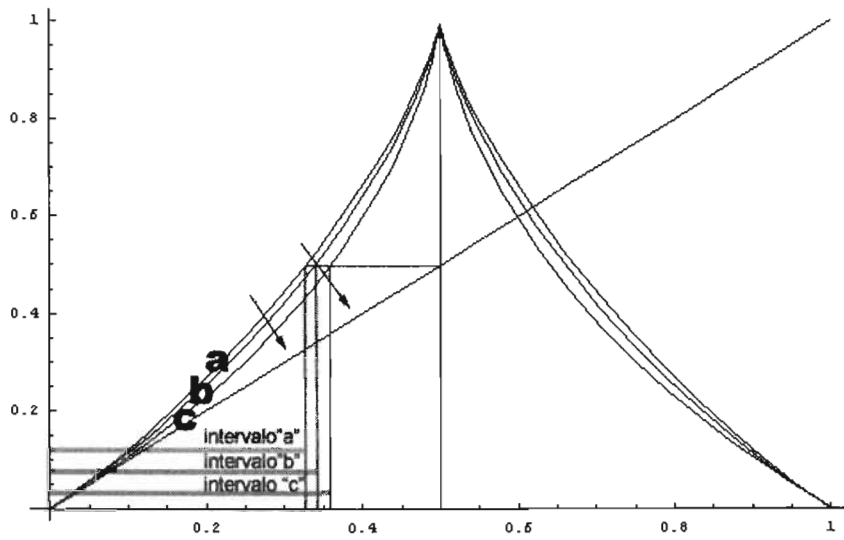


Figura 4.12: Aumento de la longitud de los intervalos de la mitad izquierda que son mapeados de nuevo a la mitad izquierda, mediante el cambio del parámetro  $\alpha$ . Los valores de los parámetro para las diferentes curvas son: para todas las curvas  $\beta = .3$ ,  $a \rightarrow \alpha = -.3$ ,  $b \rightarrow \alpha = -.6$  y  $c \rightarrow \alpha = -.9$

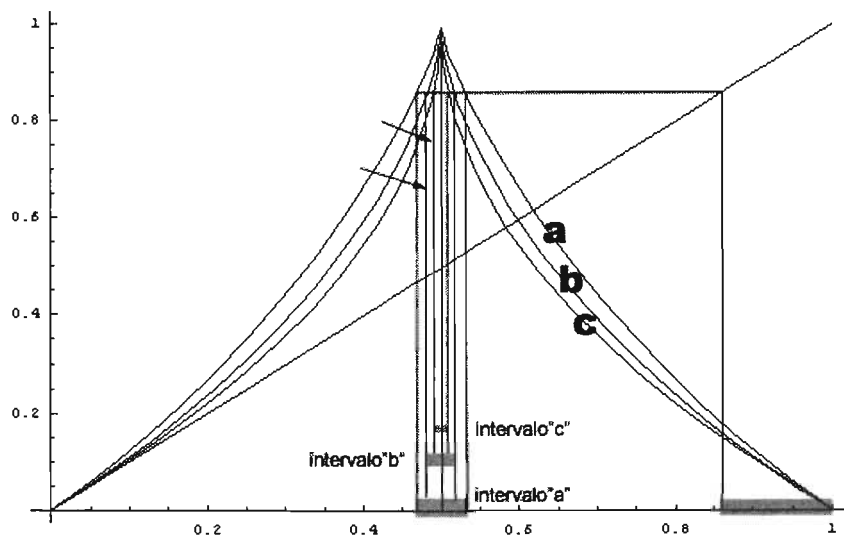


Figura 4.13: Disminución de la longitud del intervalo de la mitad izquierda que es mapeado al intervalo de la extrema derecha mediante el cambio del parámetro  $\beta$ . Los valores de los parámetro para las diferentes curvas son: para todas las curvas  $\alpha = -.3$ ,  $a \rightarrow \beta = .3$ ,  $b \rightarrow \beta = .6$  y  $c \rightarrow \beta = .9$

# Capítulo 5

## Análisis y Resultados

En los capítulos anteriores se han mostrado varios sistemas que exhiben distribuciones rango tamaño que de manera clara, como lo muestra la bondad de los ajustes, corresponden a funciones Beta de dos parámetros (con variables normalizadas). Si hacemos una comparación de los distintos mecanismos que llevan a cada uno de estos sistemas a presentar esta distribución encontraremos que existen fenómenos en ellos que pueden ser homologados de un sistema a otro, y que la manipulación de éstos tiene efectos directos sobre los parámetros de la distribución, haciendonos sospechar que son parte de la naturaleza propia de la función Beta de dos parámetros.

### 5.1 Secuencias Genéticas y Expansión Modificación

La conexión entre las secuencias genéticas y el modelo de expansión-modificación es por construcción, ya que el modelo fue hecho explícitamente para imitar los principales mecanismos de mutación. Existe evidencia clara de que los cambios en las secuencias genéticas se dan a nivel de pares de bases o de grupos numerosos de ellos pero no particularmente en codones. El hecho de que en primera aproximación el número de aminoácidos que aparecen en una secuencia suficientemente grande está relacionado directamente con el número de degeneraciones en el código que transcriben ese aminoácido, es decir, si un aminoácido en particular es codificado por cuatro codones diferentes, entonces es más probable que aparezca que un aminoácido que corresponde solamente dos o tres codones diferentes.



Figura 5.1: Frecuencia de aminoácidos en proteínas con relación a la frecuencia predicha a partir de la degeneración del código

Dada esta independencia de las mutaciones y los marcos de lectura podemos inferir que un sistema que imite los mecanismos de mutación más comunes, aunque sea una secuencia en donde no existan asimetrías como las del marco de lectura natural, tendrá una distribución del mismo tipo que las secuencias genéticas reales. Éste es el caso del modelo de expansión-modificación.

Un modelo que solo incluya uno de los dos factores de cambio, expansión o modificación, produciría distribuciones uniformes o gaussianas. Por ejemplo, consideramos una secuencia binaria aleatoria, arbitrariamente larga, y la sometemos a un proceso que incluya solamente modificación de símbolos individuales con probabilidad  $p$ , entonces el resultado es una secuencia con una distribución idéntica a la de la secuencia original, una distribución uniforme. Este hecho se hace evidente cuando pensamos de la siguiente manera: si partimos de una secuencia binaria aleatoria suficientemente grande obtendremos distribuciones rango tamaño parecidas a la uniforme, si ahora invertimos los valores individuales ( $1 \rightarrow 0$  ó  $0 \rightarrow 1$ ) de todos los elementos obtendremos, por supuesto, una distribución idéntica a la original. Si en vez de cambiar todos los símbolos solo los modificamos con probabilidad  $p$  entonces tenemos, después del proceso, dos subconjuntos de nuestra secuencia aleatoria original, uno que quedo igual y otro cuyos elementos cambiarion por su valor opuesto, para una secuencia lo suficientemente grande y un valor de  $p$  suficientemente alejado de  $1^1$  el primer conjunto tendría una distribución parecida a la original, y el segundo tendría la distribución de una serie aleatoria cuyos elementos en su totalidad fueron cambiados de valor, es decir, como vimos arriba, tendría también una distribución homogénea.

Vemos entonces que la aparición de una distribución Beta en el modelo expansión modificación (Figs. 4.8,4.9 y 4.10) es consecuencia principalmente del factor de expansión, lo que nos da una primera idea de cuáles son los

---

<sup>1</sup>Si el conjunto que no cambió sigue siendo considerablemente grande entonces es una muestra representativa de todo el conjunto, dada la aleatoriedad. El tamaño de dicho subconjunto depende no solo de el tamaño original de la muestra sino también del valor de  $p$ , que de alguna manera representa las proporciones entre la serie original y su subconjunto que no cambió

fenómenos que podrían existir en los otros modelos en donde aparece la distribución Beta.

## 5.2 El mapeo logístico y el modelo de expansión-modificación

En la sección 4.2 vimos como las secuencias binarias generadas con el mapeo logístico, agrupadas en  $n$ -ómeros, tienen una distribución Beta para valores del parámetro  $\mu$  cercanos al límite ergódico  $\mu = 4$ . Con el antecedente del modelo de expansión modificación podemos empezar a buscar un símil de la expansión del MEM en la manera en que los iterados del mapeo logístico alternan su posición en las dos mitades del intervalo  $[0, 1]$ .

En las secuencias generadas por la curva logística un equivalente al mecanismo de expansión del MEM tendría que hacer que los iterados tuvieran una tendencia a permanecer de un solo lado del intervalo por un cierto número de iteraciones para después cambiar su posición a la otra mitad y ser susceptibles de repetir el proceso. El fenómeno de *intermitencia* o al menos su ingrediente principal, la bifurcación tangencial inversa, imita este tipo de comportamiento.

### 5.2.1 Intermitencia y bifurcación tangencial inversa

El fenómeno de la intermitencia se presenta como la interrupción súbita y efímera de un comportamiento aparentemente periódico.

El principal ingrediente de la intermitencia es la bifurcación tangencial inversa, la ruta al caos más común después de la ruta por bifurcación. La bifurcación tangencial se presenta cuando se crea un punto de silla (*saddle point*) mediante la fusión de dos puntos fijos, uno estable y uno inestable, con el cambio del parámetro, partiendo la región local en una sección atractiva y en otra repulsiva.

Antes de la aparición del punto de silla existe un cambio gradual en el comportamiento local. La vecindad del punto silla, con valores del parámetro ligeramente menores al valor crítico  $\mu_c$  de la formación del punto de silla, forma un "canal" entre la curva del mapeo y la identidad, cuando el valor del parámetro se acerca a  $\mu_c$  el canal se angosta, de manera que los iterados que cruzan este canal tardan más tiempo en salir de una vecindad dada alrededor del valor del punto silla.

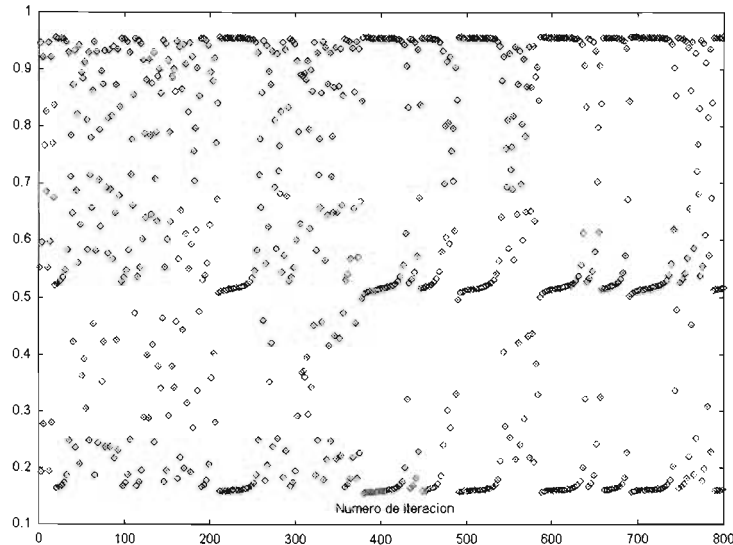


Figura 5.2: Intermittencia de una aparente órbita de periodo tres (en  $\mu = 3.828$ )

Conforme se varía el parámetro del mapeo logístico tendiendo hacia el valor de ergodicidad la densidad de ocurrencia de bifurcaciones tangenciales en las primeras composiciones ( $f \circ f = f^{(2)}, f^{(3)}, f^{(4)}, \dots$ ) aumenta ([9, 15]) y por lo tanto deberíamos apreciar una creciente tendencia de los iterados a permanecer en una mitad.

Para corroborar esta conjetura se hizo un diagrama especial para el desarrollo temporal de los iterados. El diagrama funciona como un caminante aleatorio, en donde el eje horizontal representa el tiempo y el vertical es una medida del cambio: si un iterado cambia de símbolo (región) entonces el caminante sube una unidad, en caso contrario, cuando el siguiente iterado es el mismo símbolo, el valor vertical baja una unidad. De esta manera comprobamos que en efecto hay una tendencia clara en las secuencias a aumentar el factor de permanencia cuando el valor del parámetro se aproxima al valor crítico de ergodicidad. Para las demás regiones tenemos un comportamiento relativamente sencillo también.

Podemos resumir el comportamiento de la proporción entre cambio ( $[0, \frac{1}{2}] \leftrightarrow [\frac{1}{2}, 1]$ ) y permanencia en todo el intervalo del parámetro  $\mu$  ( $\mu \in [0, 4]$ ):

$\mu < 3$  -En esta región la proporción de cambio es nula, dado que solo existe

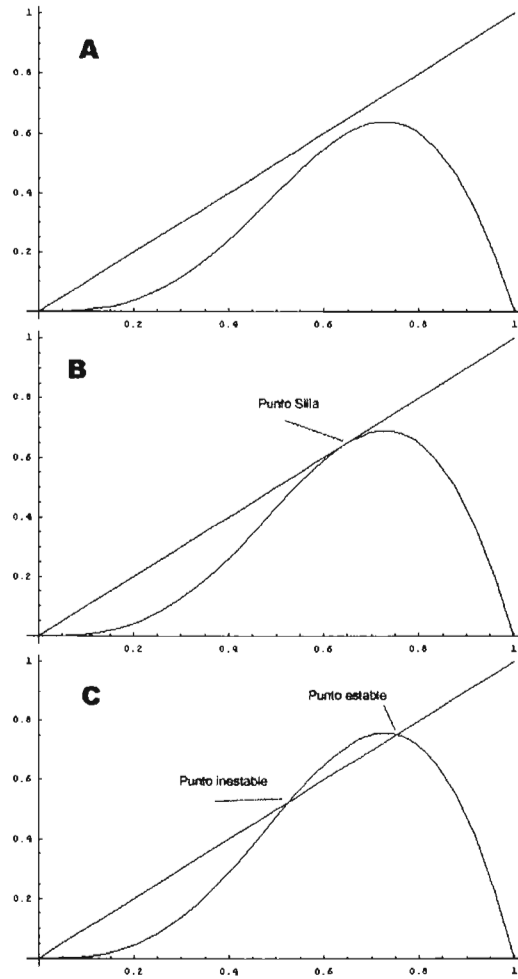


Figura 5.3: Creación de un punto de silla mediante bifurcación tangencial inversa en el mapeo  $f = ax^2 \text{sen}(\pi x)$ . La gráfica **A** muestra el mapeo con parámetro  $a = 1.6$ , en la gráfica **B** se crea el punto silla ( $a = 1.7264\dots$ ), en **C**,  $a = 1.9$ , el punto silla se separa en un punto estable y en uno inestable.

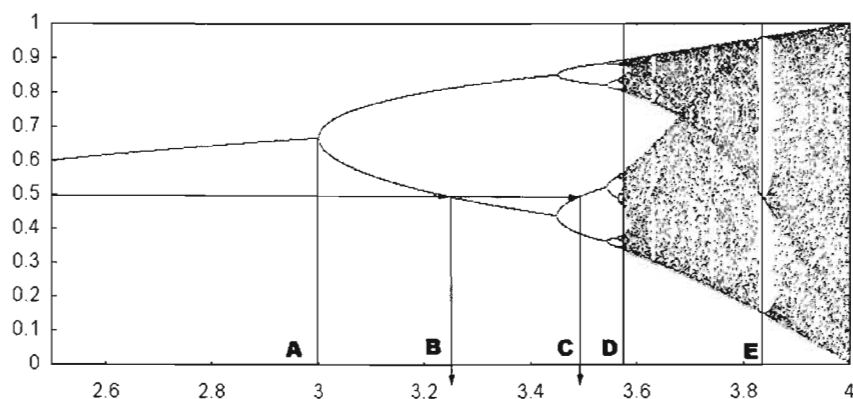


Figura 5.4: Diagrama de Feigenbaum para la curva logística. ( $D=r_\infty = 3.5699\dots$ )

un punto fijo estable de periodo 1 y eventualmente, después de un breve transitorio, el iterado caera en la mitad correspondiente para no salir de allí.

$3 < \mu < r_\infty$  -Después de la primera bifurcación ("A" en el diagrama de Feigenbaum) la proporción de cambio es nula, es hasta que el parámetro llega a un valor tal que el menor de los dos puntos de periodo 2 obtiene un valor menor a  $\frac{1}{2}$  ("B" en el diagrama de Feigenbaum) que la proporción de cambio domina totalmente el diagrama. Entre los valores "C" y "D" marcados en la figura 5.4, obtenemos secuencias del tipo  $\overbrace{1110} \overbrace{1110} \overbrace{1110}$ , por lo tanto existe tanta permanencia como cambio, en el diagrama esto se representa como una línea horizontal.

$r_\infty < \mu < 4$  -Inmediatamente dentro de la región caótica la situación cambia drásticamente, ahora la tendencia es claramente una preferencia del iterado a cambiar de mitad. En esta región la proporción entre cambio y permanencia tiende a un valor fijo, para un número moderadamente grande de iteraciones (después de 1000 iteraciones el diagrama de cambio adopta la forma de una línea. Las fluctuaciones alrededor de esta línea se pierden totalmente de escala cuando uno considera mas de 200,000 iteraciones). Cuando se aumenta el parámetro, evitando los valores de las ventanas periódicas, notamos una disminución de la proporción cambio permanencia, indicando que la notoria preferencia de

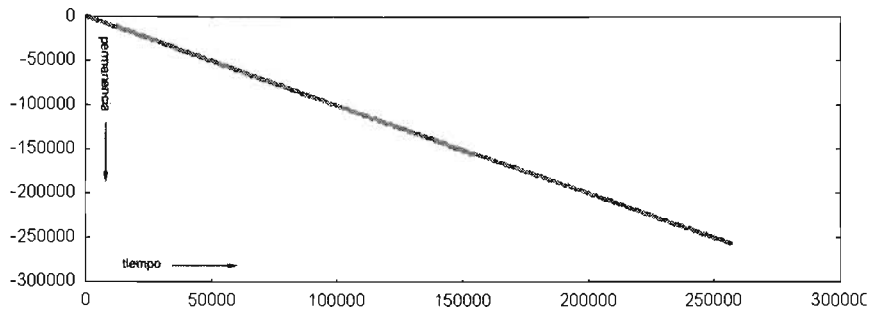


Figura 5.5: Diagrama de cambio contra tiempo. para  $\mu < 3$  El eje horizontal representa el número de la iteración, la curva sube una unidad si el iterado anterior está en la mitad en la que no está el presente iterado y baja una unidad cuando ambos iterados se encuentran en la misma mitad

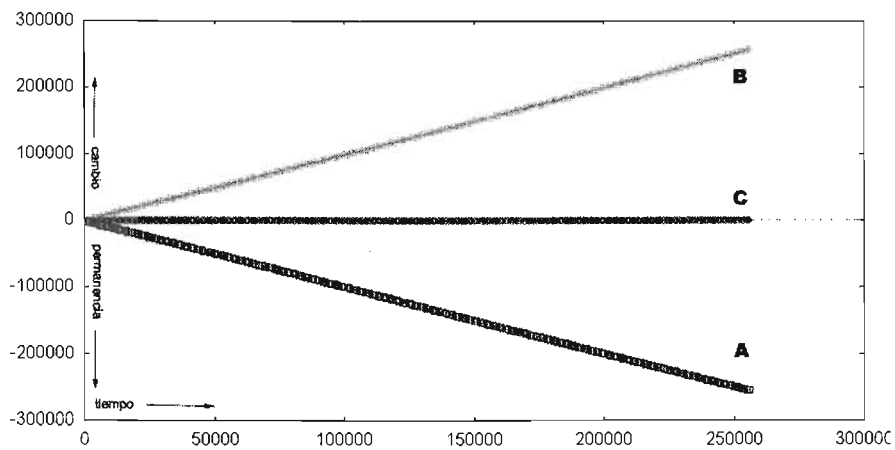


Figura 5.6: Diagrama de cambio contra tiempo. para  $3 < \mu < r_\infty$ . La curva "A" corresponde a  $\mu = 3.23 < B$ , la curva "B" corresponde a  $B < \mu = 3.38 < C$  y la curva "C" corresponde a  $C < \mu = 3.5 < D$ .

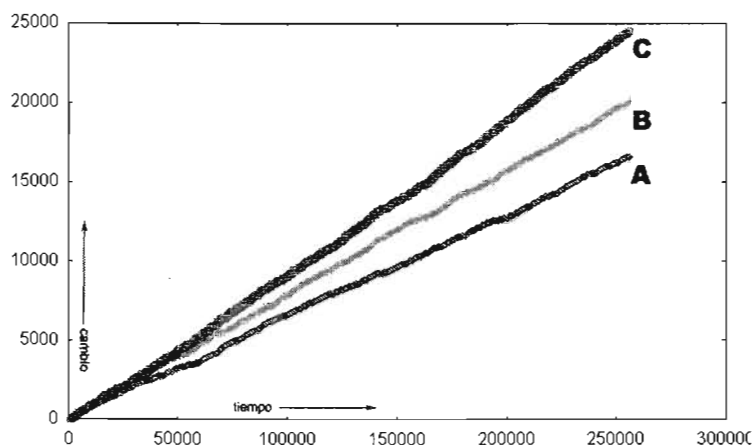


Figura 5.7: Diagrama de cambio contra tiempo para  $r_\infty < \mu < 4$ , La curva "A" corresponde a  $\mu = 3.936$ , la curva "B" corresponde a  $\mu = 3.932$  y la curva "C" corresponde a  $\mu = 3.93$ .

un iterado a cambiar de mitad disminuye visiblemente con el aumento del parámetro.

$\mu \approx 4$  -Cuando la diferencia entre el parámetro y el valor crítico 4 es del orden de milésimas, aumentan las fluctuaciones en las proporciones de cambio-permanencia, pero aún tienen un promedio claro. El aumento del parámetro hacia el valor crítico sigue teniendo como consecuencia la disminución en proporción de los cambios. Justo en el valor crítico la proporción de cambio es la misma que de permanencia, en el diagrama de cambio-tiempo observamos una recta horizontal con fluctuaciones. La pendiente en el diagrama disminuye de manera abrupta cuando el valor final del parámetro es 4.

Es con estos últimos valores del parámetro ( $\mu \approx 4$ ) que encontramos funciones Beta en la distribución rango-tamaño de las secuencias binarias, con valores menores ya vimos que el número de n-adas diferentes es muy pequeño, el equilibrio entre cambio y permanencia que obtiene el mapeo en valores altos es el responsable de darle a las secuencias binarias la diversidad suficiente como para que todos los n-ómeros posibles aparezcan alguna vez. Tenemos entonces que una vez que la proporción entre cambio y permanencia hace un salto súbito para  $\mu \approx 4$  aparece la función Beta y que el factor de

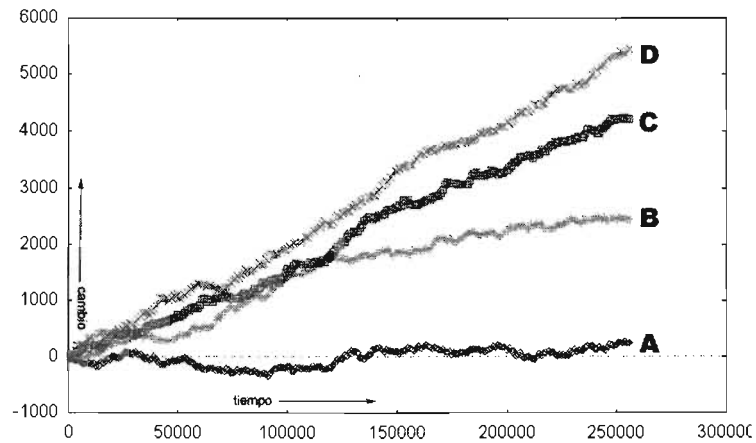


Figura 5.8: Diagrama de cambio contra tiempo. para  $\mu \approx 4$ . La curva "A" corresponde a  $\mu = 4$ , la curva "B" corresponde a  $\mu = 3.9998$ , la curva "C" corresponde a  $\mu = 3.9996$  y la curva "D" corresponde a  $\mu = 3.9992$ .

expansión en el MEM es responsable de la forma de la función Beta en su distribución, ligando, como esperabamos, el factor de expansión del MEM con el mecanismo de permanencia en el mapeo logístico con  $\mu \approx 4$ .



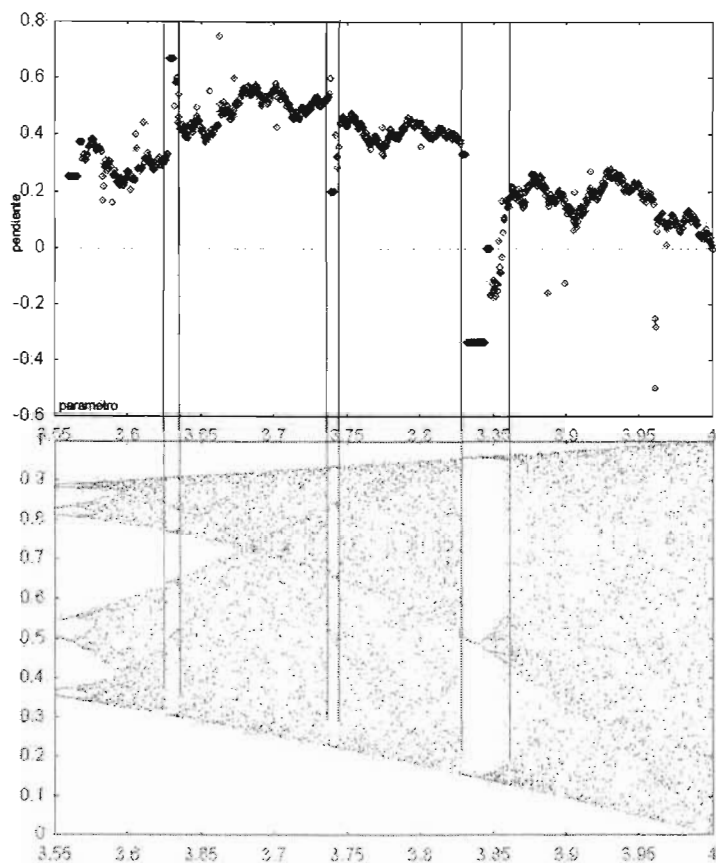


Figura 5.9: Pendiente del diagrama de cambio-permanencia contra  $\mu$ . Debajo se muestra el diagrama de Feigenbaum, para localizar las ventanas periódicas principales e identificarlas con las interrupciones en la disminución de la proporción de cambio.

# Capítulo 6

## Naturaleza de *TPBF* (Conclusiones)

### 6.1 La distribución rango-tamaño en secuencias de Codones.

Las secuencias de n-ómeros ( y en particular en codones) en ADN procarionte y en exones eucariontes tienen como distribución rango-tamaño (r-t) una función Beta de dos parámetros. El hecho de que esto suceda en n-ómeros y no solo en codones hace pensar que las mutaciones en efecto se dan a nivel de bases, no como un código explícitamente cambiado para optimizar la sintetización de proteínas. La forma compleja de la distribución r-t de las secuencias completas de ADN eucarionte se debe al carácter inhomogéneo de su estructura, una vez que se hace el estudio en secciones codificadoras (en un gen completo o en exones) se obtienen los resultados que aparecen en las secuencias procariontes. La inhomogeneidad. Es posible que la función Beta también se ajuste a las distribuciones r-t de las regiones no codificadoras, si de ellas excluimos las regiones de repetición que generan los transposones (en especial los retrotransposones de la familia ALU). Los retrotransposones ALU son cadenas de alrededor de 300 pb de largo que tienen la facultad de hacerse copiar por ARNm para después ser introducido en cualquier otro lugar dentro de la secuencia. En el genoma humano se han encontrado miles de millones de repeticiones de estos retrotransposones, de manera que ahora su masa representa alrededor del 11% de la masa total del código. Este mecanismo debe contribuir de manera fuerte a la inhomogeneidad de los códigos

eucariontes. Se puede concluir que las mutaciones puntuales en el código genético proveen una condición suficiente para la aparición de la función Beta de dos parámetros en la distribución r-t, dado que la imitación de estos mecanismos via el modelo E-M genera distribuciones Beta. Los parámetros de la distribución Beta r-t de un organismo son propios de la especie, lo cual provee un potencial método de clasificación genético. Se propone mapear los parámetros de diferentes especies en un plano  $\alpha - \beta$  e investigar si existe algún tipo de continuidad por regiones. Dado que los parámetros de la función Beta son cantidades de fácil medición, un mapeo como el descrito podía llegar a proveer una método sencillo y rápido de categorización filogenética para muestras grandes de código genético.

## 6.2 Distribución rango-tamaño en secuencias binarias del mapeo logístico.

Las secuencias binarias generadas por el redondeo de los iterados del mapeo logístico en el régimen caótico, cerca del valor de ergodicidad, tienen por distribución r-t una función Beta. El factor expansión modificación está correlacionado con la aparición de la función Beta. El factor de expansión-modificación presenta una gran sensibilidad ante los valores del parámetro cercanos al de caos total ( $\mu = 3.999 \pm 0.001$ ), y es solo en éste rango de valores donde la distribución r-t es una función Beta. El mapeo logístico es incapaz de producir todos los n-ómeros diferentes si el parámetro se encuentra ligeramente por debajo del rango antes descrito. Las repeticiones de secuencias de "n" símbolos se debe a la presencia de una bifurcación tangencial (normal o inversa) en la n-ava composición del mapeo.

## 6.3 TPBF como distribución rango-tamaño

Cuando la función Beta de dos parámetros representa una distribución rango-frecuencia (rango-tamaño, en donde la cantidad medida es la frecuencia de aparición) podemos interpretar los parámetros  $\alpha$  y  $\beta$  como estimadores de que tantos valores tienen frecuencias muy por debajo o muy por encima de una frecuencia promedio. Un valor de  $\beta$  próximo a cero y menor a la unidad, indicaría, por ejemplo, que existen pocos valores (n-ómeros) con frecuencias por debajo del promedio. De manera parecida, tenemos que un valor negativo

de  $\alpha$  cercano a cero refleja la escasez de n-ómeros con frecuencias por encima del promedio.

## 6.4 Modelos generales generativos de TPBF como distribución rango-tamaño.

La existencia de mecanismos tipo expansión-modificación es una condición *suficiente* para la aparición de una función Beta como distribución rango tamaño. Para encontrar una condición más general o quizás una necesaria y suficiente se ha propuesto, fuera del trabajo de esta Tesis, que la degeneración propia de algún producto específico de distribuciones aleatorias puede generar una distribución Beta como límite. Por ejemplo, suponganse dos colecciones discretas de  $N$  valores, la colección de todas las  $N^2$  sumas (o multiplicaciones) posibles, organizada como rango tamaño, produce curvas *sigmoides* que se ajustan bien con una función Beta. Si llevamos el ejemplo al continuo, hayar la distribución r-t de la "suma" de varias distribuciones r-t continuas. Se hicieron pruebas numéricas para obtener la distribución rango-tamaño obtenida de la suma de  $n$  distribuciones lineales continuas, los resultados se ajustaban bien a un TPBF. Quizás sea posible demostrar qu el límite cuando  $n \rightarrow \infty$  de la distribución de la "suma" de  $n$  distribuciones lineales es una función Beta. Se obtuvo ya una expresión analítica para el inverso de la distribución r-t final que resulta de sumar  $n$  distribuciones lineales, solo es necesario encontrar el límite al infinito o demostrar que la diferencia con una distribución TPBF tiende a cero. De encontrarse alguna relación entre la "suma" de una forma genérica de funciones crecientes y la distribución Beta podríamos proponer que la forma más general de explicar la aparición de una TPBF como distribución r-t, sería la de un mecanismo que dependiera del producto de varias distribuciones independientes, y que a manera convergencia de límite central, los parámetros  $\alpha$  y  $\beta$  de su distribución convergieran hacia dos valores definidos.

# Capítulo 7

## Análisis de Universalidad de *TPBF*

Con el análisis de esta tesis concluimos que los mecanismos tipo expansión modificación que presentan tanto el mapeo logístico como la dinámica de modificación de las secuencias genéticas producen en las distribuciones r-t una función Beta. Para hacer conclusiones acerca de que tan generales son estas condiciones podemos primero analizar otros sistemas cuyas distribuciones r-t se ajustan con una distribución Beta.

### 7.1 Series Financieras

En [18] se hace el análisis de las caídas porcentuales de el índice Standard& Poor 500 en la bolsa de valores de Nueva York, con los datos registrados cada minuto, durante un año. Presentan varios modelos para explicar la relación entre la frecuencia de aparición de un cambio en la bolsa con respecto al tamaño porcentual de éste, entre los cuales destacan un modelo con distribución natural, y otro con una distribución estable de Levy. La diferencia principal de los ajustes hecho con una u otra distribución residen en el rango de validéz. El modelo de vuelo de Levy tiene un ajuste excelente para los eventos de alta probabilidad (alta frecuencia), pero falla por completo en ajustar los eventos menos frecuentes, que corresponden en general a caídas del índice más grandes (ver fig 7.1).

Un ajuste con *TPBF* a la distribución resulta exitoso (RMS=0.995) en TODO el rango de caídas que se presenta en el trabajo, lo que sugiere que

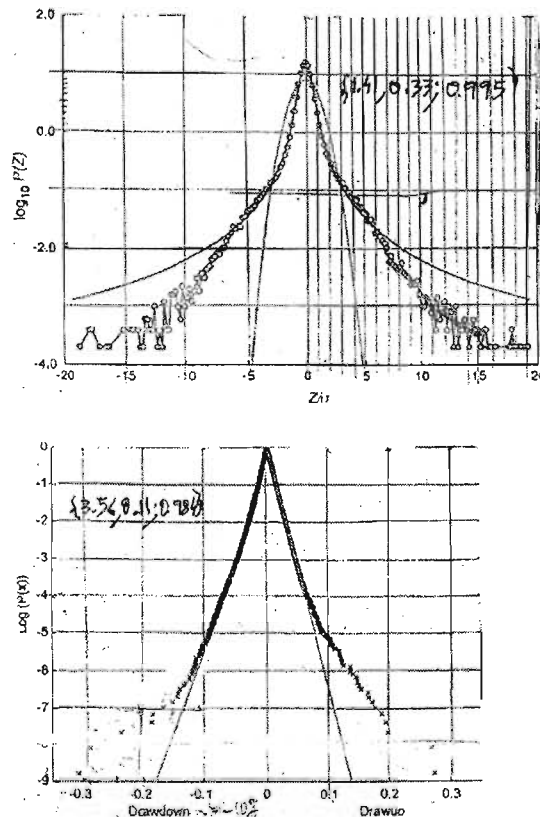


Figura 7.1: Gráfica superior: curva del cambio en el logaritmo de dos valores consecutivos (en la línea temporal) del índice S&P 500 contra  $\log_{10}$  de la probabilidad, el ajuste central es una curva Gaussiana, el ajuste en todo el rango es el ajuste con modelo estable de Levy. Gráfica inferior: Caídas porcentuales del mismo índice contra  $\log_{10}$  de la probabilidad. En manuscrito se presentan los valores de  $-\alpha, \beta$  y el RMS respectivamente del ajuste de una TPBF.

se pueden unificar la dinámica que causa los cambios en la bolsa, es decir, que un modelo que procure que las distribuciones  $r-t$  sean funciones Beta no necesitaria tomar por separado los casos de caídas grandes y pequeñas, como resulta cuando adoptamos los modelos propuestos hasta ahora.

No es evidente que los cambios en los índices económicos se puedan explicar con un modelo sencillo tipo expansión modificación, de manera que la aparición de una función Beta en la relación probabilidad-tamaño de la caída parece indicarnos que los mecanismos de expansión-modificación no son condiciones necesarias para obtener una *TPBF*, y que la dinámica esencial atrás de los cambios en la bolsa de valores obedece a una serie de condiciones mucho más generales.

## 7.2 *TPBF* en Música Armónica

En el primer capítulo mencionamos como la relación entre la frecuencia y la intensidad de una señal sonora musical sigue, en ciertos casos, una ley de Zipf. Si ahora nos restringimos a estudiar la estructura armónica de la música encontramos distribuciones parecidas, pero de una naturaleza bastante diferente.

Si hacemos un conteo de las veces que se utilizó cada una de las diferentes notas en una pieza musical, y hacemos una distribución rango tamaño, encontramos que ésta última se ajusta muy bien por una *TPBF*, mucho mejor, por ejemplo, que el ajuste con ley de Zipf que se adopta comúnmente ([19]). Se hicieron trabajos sobre música de Beethoven, Bach, Mozart, Debussy y Schoenberg, y en todos se corroboró la fidelidad de la función Beta. En la figura 7.2 se muestra un resultado para una obra de Beethoven.

En la estructura armónica, cada nota tiene una función específica dado un marco armónico inicial llamado *clave*, y aquí, por nota, nos referimos a una nota (Do, Sol, Fa#, etc...) y a todas sus octavas (Do1, Do2, Do3, etc...), i.e., a todas las notas que resulten de duplicar sucesivamente la frecuencia asociada a la nota original. Resulta interesante entonces hacer la distribución  $r-t$  de las notas junto con sus octavas, para darle un sentido más armónico. Los resultados de este nuevo análisis también tuvieron un buen ajuste con función Beta, pero con un valor RMS ligeramente menor ( $\approx .01$ , ver fig 7.3 y 7.4).

Los resultados anteriores, en combinación con el conocimiento de las reglas elementales de armonía, sugieren que la distribución Beta en el conjunto

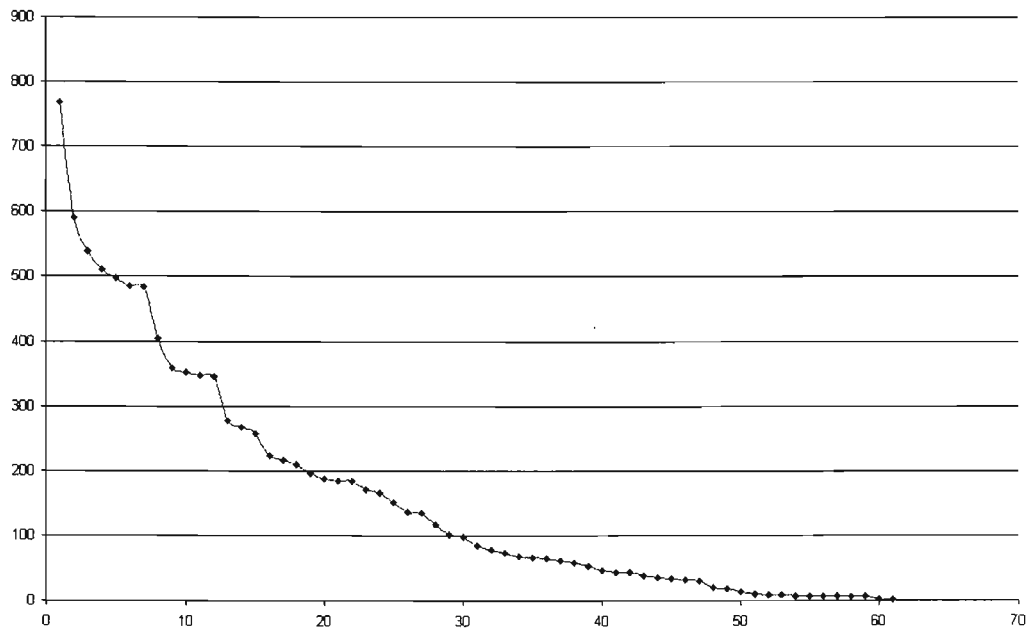


Figura 7.2: Distribución rango-tamaño de las notas (diferenciando las mismas notas de diferente octava) de la Sonata No.25, Op.79 para piano de L.V. Beethoven. El ajuste resultó tener:  $RMS > 0.99$



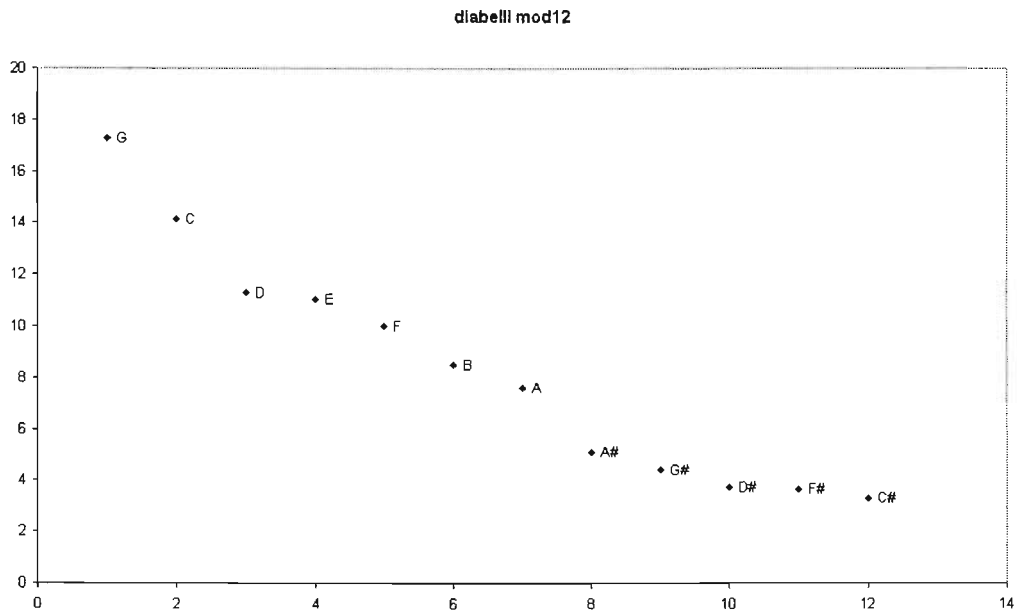


Figura 7.3: Distribución rango-tamaño de las notas (tomando diferentes octavas como la misma nota) del conjunto de 33 variaciones en Do mayor sobre un tema de A. Diabelli de L.V. Beethoven. Las cinco primeras notas corresponden a las notas de la escala armónica principal de la tonalidad dominante en el conjunto de variaciones.

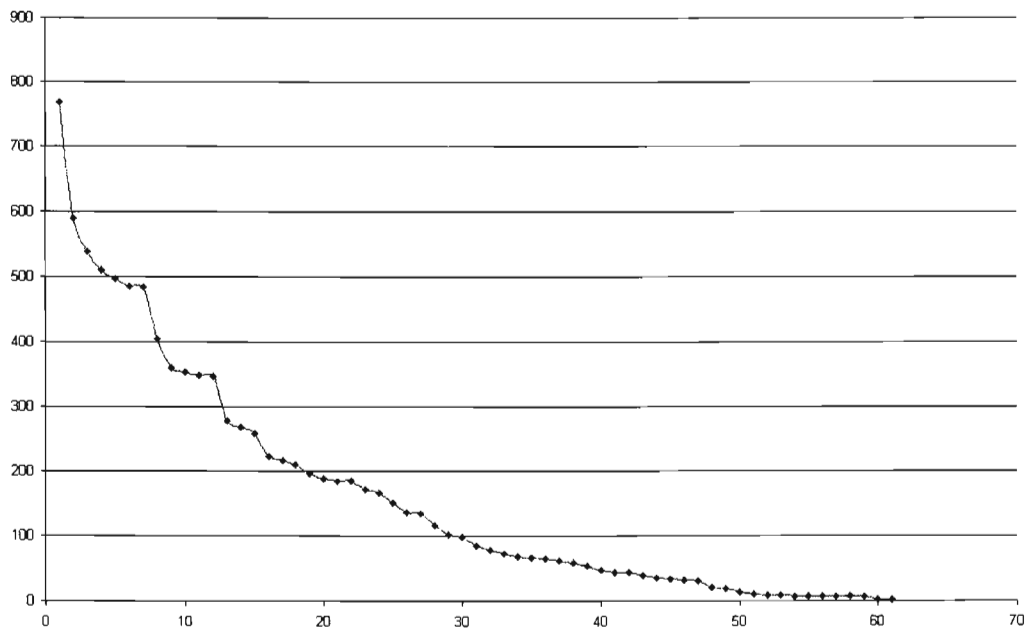


Figura 7.4: Distribución rango-tamaño de las notas de la Sonata para Cello en G de C.A. Debussy

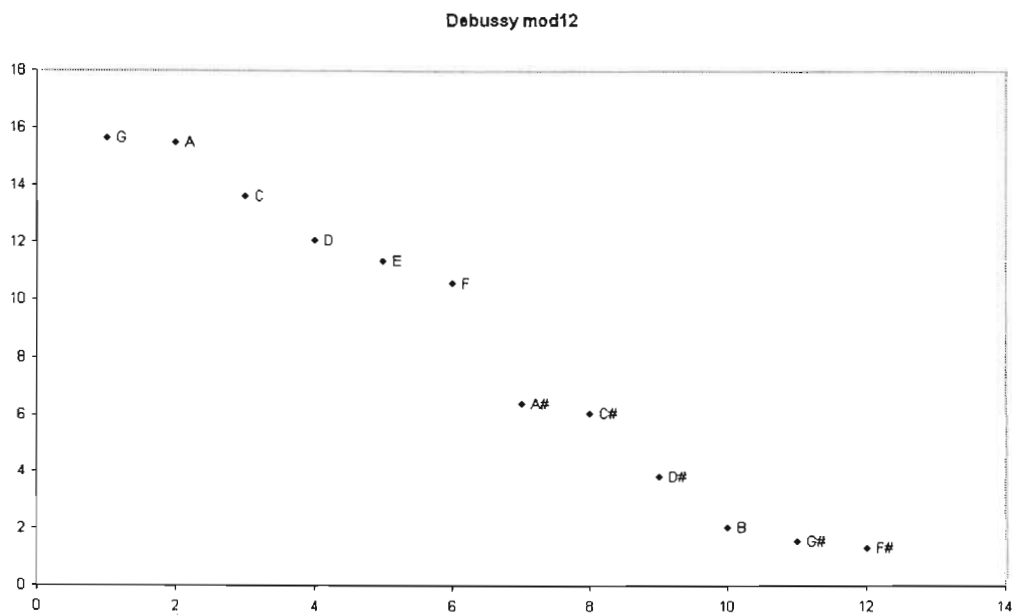


Figura 7.5: Distribución rango-tamaño de las notas (tomando diferentes octavas como la misma nota) del conjunto de 33 variaciones en Do mayor sobre un tema de A. Diabelli de L.V. Beethoven. El ajuste en rojo es el de una función Beta:  $\alpha = -0.3$ ,  $\beta = 1.1$ ,  $RMS = 0.97$

de notas con sus octavas es consecuencia de la dependencia tipo Markov a primer vecino de una nota dada su antecesora. En [16] vemos que un mecanismo generador de leyes de potencia es el llamado "*Preferential Attachment*", que es exactamente lo que pareciera dominar sobre la secuencia de notas en una composición armónica.

Tenemos entonces que un modelo de Markov a primer orden de las notas de la escala musical con reglas armónicas arbitrarias (los compositores cuyos trabajos fueron analizados se escogieron de tal manera que las reglas armónicas de uno y otro fueran lo más diferentes posibles) es capaz de producir funciones Beta, y que una re-distribución de dichas notas con un mecanismo menos riguroso (más aleatorio) sobre las octavas posibles mejora aún más el ajuste con una función Beta. Vemos pues que en la música tenemos una combinación de dos dinámicas que contribuyen independientemente a la aparición de una función Beta.

# Bibliografía

- [1] Schroeder, Manfred "Fractals, Chaos, Power Laws: Minutes from an Infinite Paradise" Ed. Freeman & Company, Septiembre 1992
- [2] Reed, William "The Pareto, Zipf and other power laws" C49;D31;R12
- [3] Reed, William "On the Rank-Size Distribution for Human Settlements" R12;C49
- [4] Marc D. Hauser, Noam Chomsky, W. Tecumseh Fitch "The Faculty of Language: What Is It, Who Has It, and How Did It Evolve?" SCIENCE VOL 298,22 NOVEMBER 2002
- [5] Ferrer i Cancho, Ramon-Sole, Ricard V. "Least effort and the origins of scaling in human language" 788,791 PNAS Febrero 4, 2003 vol. 100 no. 3
- [6] Patel, Anniruddh "Language, music, syntax and the brain" Nature, Vol.6 No.7 pp 674-681 July 2003
- [7] Koga, Shinji "The inverse Problem of Frobenius Perron Equations in 1D Difference Systems" Progress of Theoretical Physics, Vol. 86, No5, November 1991
- [8] Koga, Shinji "Phase Description Method to Invariant Densities in 1-D Difference Systems" Progress of Theoretical Physics, Vol. 80, No2, August 1988
- [9] Schuster, Heinz Georg "Deterministic Chaos: An introduction" Ed. VCH Verlagsgesellschaft mbH.

- [10] D.Pingel, P.Schmelcher, F.K.Diakonos "Theory and examples of inverse Frobenius-Perron problem for complete chaotic maps" Chaos, Vol-9 No.2, Junio 1999
- [11] K. Arai, Y. Terayama, T. Arata "Image Classification Based on Beta Distribution for SAR Image"
- [12] Lewin,Benjamin. "Genes" Prentice Hall
- [13] A. Baranovsky, D.Daems "Design of one-Dimensional Chaotic Maps with Prescribed Statistical Properties" International Journal of Bifurcation and Chaos Vol.5, No.6 1995
- [14] N.Metropolis, M.L. Stein, P.R. Stein "On Finite Limit Sets for Transformations on the Unit Interval" Journal of Combinatorial Theory. Vol.15 No.1 Julio 1973
- [15] H.Peitgen, H.Jurgens, D.Saupe "Chaos and Fractals-New Frontiers of Science" Springer-Verlag
- [16] M. Mitzenmacher "A brief history of generative models for power law and Lognormal distributions" Proceedings of the 39th annual Allerton Conference on Communication, Control and Computing,pags 182-191,2001
- [17] D.Sornette, R.Cont "Convergent multiplicative processes repelled from zero: power laws and truncated power laws"
- [18] R.N. Mantegna, H.E. Stanley "Scaling behaviour in the dynamics of an economic index" Nature Vol376 46-49 6 Julio 1995
- [19] B. Manaris, T. Pursewal, C. McCormick "Progress Towards Recognizing and Classifying Beautiful Music with Computers" IEEE 0-7803-7252-2