

01168¹²²



UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO

FACULTAD DE INGENIERÍA
DIVISIÓN DE ESTUDIOS DE POSGRADO

TESIS

GRADOS DE LIBERTAD EFECTIVOS EN ALGORITMOS GENÉTICOS

PRESENTADA POR:
ROSALÍA AGUIRRE HERNÁNDEZ

PARA OBTENER EL GRADO DE:
MAESTRO EN INGENIERÍA
(INVESTIGACIÓN DE OPERACIONES)

DIRIGIDA POR:
DR. CHRISTOPHER RHODES STEPHENS STEVENS

CIUDAD UNIVERSITARIA

MÉXICO, D.F. 1999

TESIS CON
FALLA DE ORIGEN

271506.



Universidad Nacional
Autónoma de México



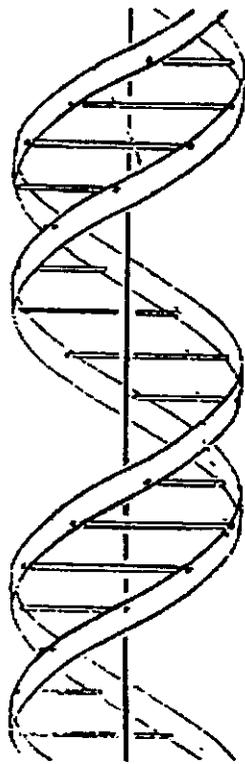
UNAM – Dirección General de Bibliotecas
Tesis Digitales
Restricciones de uso

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.

Grados de Libertad Efectivos en Algoritmos Genéticos



Rosalía Aguirre Hernández

Agradezco al Instituto de Ciencias Nucleares la beca de lugar que me otorgó para realizar la tesis y a CONACYT su apoyo económico durante los 2 años de maestría. También les agradezco al Dr. Stephens y al Dr. Michaelian su minuciosa revisión de la tesis y sus valiosos comentarios; así como al Dr. Aceves por su apoyo durante toda la maestría. Por otro lado, aprecio la ayuda que me brindó Luis para la elaboración de los programas de cómputo y la de Alberto para la edición de la tesis.

A mis padres, a mis hermanos y a Alberto.

CONTENIDO

Introducción	i
1 Preliminares.	1
1.1 Sistemas Complejos.	1
1.2 Algoritmos Genéticos.	3
1.3 Operadores Genéticos.	5
1.4 Teoría Estándar de AGs.	7
2 Transiciones de Fase.	11
2.1 Transiciones de Fase en la Física.	11
2.2 Modelo Matemático de un AG.	18
2.2.1 Ecuaciones de Evolución para Modelos Sencillos.	18
2.2.2 El Efecto de la Selección en un AG.	20
2.2.3 El Efecto de Selección y Mutación en un AG.	26
2.2.4 Ecuación de Evolución con Todos los Operadores Genéticos.	31
2.3 El Modelo de Eigen y Schuster.	33
3 Resultados Experimentales de Transiciones de Fase.	41
4 Grados de Libertad Efectivos.	57
4.1 Ejemplo de un AG con Selección y Cruza.	57
4.2 Ecuación de Evolución para Esquemas.	58
4.3 Aptitud Efectiva	69
5 Resultados Experimentales de Grados de Libertad Efectivos.	71
Conclusiones	85
A Subrutinas del Programa del AG	87
Bibliografía	91

INTRODUCCIÓN

Existen muchos problemas de optimización combinatoria que son *NP*-completos, es decir, son problemas difíciles de resolver ya que no se ha encontrado ningún algoritmo polinomial en tiempo para atacarlos. Por lo tanto, las técnicas convencionales de programación lineal o entera no son eficientes. Sin embargo, actualmente existe una tendencia por desarrollar técnicas o algoritmos que imitan algunos fenómenos naturales y que han demostrado ser más eficientes en la solución de problemas complejos [1]. Este trabajo está enfocado al estudio de los **algoritmos genéticos (AGs)**, los cuales imitan el proceso de evolución por selección natural y la genética de poblaciones.

Al implementar un AG, se debe saber qué tipo de problemas resuelve y cómo encuentra las soluciones óptimas de estos problemas. No se ha desarrollado mucha teoría que explique la manera en cómo un AG busca soluciones óptimas en un problema determinado. El teorema de esquemas [2] proporciona sólo una cota inferior sobre el número de copias de cada individuo de la población en cada generación. En este trabajo, se presenta una ecuación de evolución exacta, desarrollada por Stephens y Waelbroeck [3], a partir de la cual se obtiene el número esperado de copias de cada individuo de la población como función de las generaciones. La importancia de la ecuación de evolución radica en que a partir de ella se pueden intentar determinar cuantitativamente los grados de libertad efectivos de un AG. Es decir, las variables relevantes a partir de las cuales un AG construye soluciones óptimas. Una vez identificados los grados de libertad efectivos de un AG, se puede estudiar el comportamiento de éstos cuando se modifican los parámetros de un AG. Esto permite comprender mejor el papel que tiene cada uno de los operadores genéticos en la búsqueda del óptimo. Como consecuencia, se puede tener una idea de cómo se va a comportar un AG en un problema determinado y de los valores que deben asignarse a los parámetros del algoritmo. Por otro lado, uno de los objetivos de este trabajo es presentar los AGs en un contexto más general. Se describen los sistemas complejos, como algunos problemas de optimización combinatoria, que son el tipo de problemas que resuelve un AG.

La estructura de este trabajo es la siguiente. En el capítulo 1, se presentan conceptos preliminares sobre AGs y sobre sistemas complejos. En el capítulo

2, se desarrolla la ecuación de evolución y se estudian las características de selección y mutación a través de las transiciones de fase. Los resultados teóricos se comprueban experimentalmente en el capítulo 3. En el capítulo 4, se caracterizan los grados de libertad efectivos de un *AG* los cuales son esquemas (o bloques constructores) con determinadas propiedades. Se mostrará que las propiedades descritas en el Teorema de Bloques Constructores [2] no son totalmente correctas. En el capítulo 5, se presenta evidencia experimental que respalda estos resultados.

CAPÍTULO 1

Preliminares.

El objetivo de este capítulo es presentar un panorama general de *AGs* como modelos de sistemas complejos y como sistemas complejos en sí. Se describe la forma en cómo surgen los *AGs* al modelar sistemas complejos de la naturaleza. Se mostrará que estos algoritmos imitan algunos fenómenos de la naturaleza (a través de los operadores genéticos) y esto permite resolver problemas complejos de manera más eficiente. Por último, se estudiará la teoría estándar de *AGs* la cual describe cómo busca un *AG* soluciones aptas de un problema.

1.1 Sistemas Complejos.

El hombre se ha interesado en estudiar sistemas como el aprendizaje en los seres humanos, el sistema inmunológico, la evolución biológica, la genética de poblaciones, etc. Todos estos son ejemplos de **sistemas complejos** (*SC*). No es fácil definir un *SC*, sin embargo, la idea básica es la siguiente: entre más complejo sea un sistema, más cosas se pueden decir de él. Por ejemplo, considérense las moléculas en un cuarto vacío. Se puede hacer una descripción microscópica del sistema al mencionar la posición de cada molécula del cuarto. Sin embargo, en la práctica esto no es posible. Los **grados de libertad efectivos** de un sistema son las variables relevantes de éste con las cuales es factible describir el sistema. Dado que se ignoran algunas variables, la descripción del sistema se realiza a una escala macroscópica. En este sistema, la temperatura del cuarto proporciona una descripción macroscópica que da una idea de la energía cinética o del movimiento que tienen las moléculas. Dado que la temperatura es el único grado de libertad efectivo del sistema, éste es un sistema simple donde las velocidades de las moléculas tienen la distribución de Boltzmann. En este caso, se puede considerar que el movimiento de cada molécula es independiente de las demás. Ahora, considérese una sala de conciertos donde el público aplaude. En este caso, los aplausos provocan un movimiento en las moléculas del aire de tal manera que hay interacciones entre éstas por lo que la distribución de velocidades ya

no es la de Boltzmann. Es decir, el movimiento de una molécula depende de las demás. Sin embargo, éste tampoco es un sistema complejo porque no hay mucho que decir dado que los aplausos son aleatorios. Por otro lado, una sala de conciertos donde hay una orquesta que interpreta una obra de Beethoven, sí es un sistema complejo. Aquí, el sonido también produce un movimiento colectivo en las moléculas del aire. Sin embargo, es muy complicado explicar cómo este movimiento de las moléculas transmite la información contenida en la obra. Considérese un segundo ejemplo donde se tiene una cadena de *bits* que pueden valer 1 ó 0. La cadena 11...1 es un sistema simple ya que se describe totalmente al decir que todos los *bits* valen 1. Una cadena de *bits* aleatoria tampoco es un sistema complejo ya que se puede decir que la cadena representa los resultados del lanzamientos de una moneda honesta y de esta manera se describe totalmente el sistema. Sin embargo, la secuencia de *bits* que representa un libro tiene mucha más información como el estilo, la elección de las palabras, etc. Si el libro es realmente complejo, hay muchas cosas que decir de él. Con el objeto de entender mejor qué es un *SC*, a continuación se presentan varios conceptos relacionados entre sí que caracterizan a estos sistemas.

Los *SC* involucran fenómenos donde hay interacciones de muchos elementos que se auto-organizan en un sistema de mayor nivel y por lo tanto muestran emergencia de estructura. Por ejemplo, la sincronización de una bandada de aves emerge de las interacciones entre éstas pero sin un mecanismo externo que los controle. Otro ejemplo es el sistema inmunológico de los mamíferos el cual está compuesto por una colección de células especializadas que se auto-organizan para realizar determinadas tareas.

Otra característica que presentan los sistemas complejos es la **frustración**. Es común que en la vida se tengan metas incompatibles. La única solución es abandonar algunas de estas metas. Pero al decidir cuál abandonar, uno se siente frustrado. Por ejemplo, es difícil ser amigo de 2 personas que se odian. Se debe elegir una de ellas y esto es una situación frustrante. Un problema típico de optimización donde hay frustración es el problema del agente viajero. Considérese un conjunto de ciudades y la distancia que hay entre éstas. El problema consiste en encontrar el orden en que deben visitarse las ciudades para recorrer la menor distancia posible. El problema presenta frustración ya que la solución óptima no siempre consiste en dirigirse a la ciudad más cercana que no se haya visitado. De hacerlo, la longitud total del recorrido sería mayor que la solución óptima.

La presencia de frustración implica que hay muchos óptimos locales en la función que se desea optimizar. Si se realiza una representación gráfica, en el espacio de las variables del problema, de la aptitud de los diferentes estados del sistema (mínimos, máximos, puntos silla) se obtiene el **paisaje de aptitud**. Cabe hacer notar que las características del paisaje de aptitud de un sistema dependen de la métrica que se use. Sin embargo, en los casos donde hay muchos

óptimos locales, se obtienen **paisajes de aptitud rugosos** con muchos valles y montañas. Por lo tanto, los métodos de máximo descenso generalmente no encuentran el óptimo global ya que realizan la búsqueda al pasar de una solución a otra mejor. Además, existen sistemas adaptivos donde el paisaje de aptitud se modifica con el tiempo, es decir, no es estático. Por ejemplo, en el caso del problema del agente viajero, el tiempo o el costo para ir de una ciudad a otra se puede modificar con el tiempo. En consecuencia, el sistema no contiene un óptimo global por lo que la optimización no es tan importante como el mejoramiento de las soluciones.

Las características presentadas anteriormente describen de manera general los sistemas complejos. Sin embargo, para estudiar el comportamiento de estos sistemas, se realizan modelos artificiales como los *AGs*. Por ejemplo, se puede modelar una población del virus de inmunodeficiencia humana (VIH) en un paciente a través de un *AG* y estudiar el papel que tienen las mutaciones en el VIH para escapar del sistema inmunológico. Por otro lado, los *AGs* imitan algunos fenómenos de la naturaleza para resolver problemas de optimización complejos aprovechando de esta manera el éxito que ha tenido la naturaleza en la preservación y adaptación de los individuos más aptos.

1.2 Algoritmos Genéticos.

Los *AGs*, desarrollados por John Holland a mediados de los 60, son procedimientos de búsqueda basados en mecanismos de selección natural y genética de poblaciones. En la naturaleza, los individuos de cualquier especie están sometidos a una lucha continua por sobrevivir. Los individuos más aptos tienen mayores probabilidades de sobrevivir, reproducirse y transmitir a sus descendientes las características que los hacen superiores al resto de la población, mientras que los individuos menos capaces forman una minoría o desaparecen. A través de la selección natural, la naturaleza favorece a los individuos que tienen determinadas características. Estas características se determinan a nivel genético por la manera en que los cromosomas de los padres se combinan, por ejemplo en la reproducción sexual. De manera similar, los *AGs* combinan pedazos de soluciones existentes para encontrar mejores soluciones en problemas complejos.

Los *AGs* difieren de las técnicas convencionales de optimización en varios aspectos. Por un lado, los *AGs* realizan una búsqueda en paralelo ya que trabajan con una población de individuos (es decir, con un conjunto de soluciones) que se conocen como **cuerdas** o **cromosomas**. Cada cromosoma tiene asociado una **aptitud** que está determinada por el valor de la función objetivo de la solución correspondiente.

Otra característica de los *AGs* es que los cromosomas constituyen una codificación de las soluciones del problema bajo consideración. En forma natu-

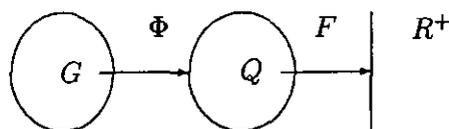


Figura 1.1: Codificación indirecta.

ral, la función de aptitud se aplica directamente sobre los individuos físicos (**fenotipos**). Sin embargo, al elaborar un modelo se trabaja con las características de los individuos las cuales se codifican en una cadena de caracteres (**genotipo**). Por ejemplo, un fenotipo puede ser el sexo de un individuo, mientras su genotipo es la parte correspondiente al par 23 de su cromosoma (xx para la mujer y xy para el hombre). La codificación se puede hacer de distintas formas. En el método de codificación indirecta, las variables del problema se representan en los genes del cromosoma mediante un intérprete que traduce los valores de un cromosoma en valores de la variable del problema. Entonces, la relación entre la codificación y la función de aptitud se puede observar en la figura 1.1. Los genotipos (elementos de G) son asociados con los fenotipos (elementos de Q) mediante la codificación Φ . Estos fenotipos son evaluados por la función de aptitud F para obtener un valor real positivo. La codificación binaria y la de Gray son ejemplos de codificaciones indirectas. En la codificación de Gray, los reales x (fenotipos) que difieren en una unidad, se codifican de tal manera que sólo difieren en el valor de uno de sus *bits*. En la siguiente tabla se ejemplifican las 2 codificaciones.

x	codificación binaria	codificación de Gray
0	000	000
1	001	001
2	010	011
3	100	110

Los AGs son métodos de búsqueda global ya que usan reglas de transición probabilísticas para guiar la búsqueda hacia regiones donde la aptitud de las soluciones sea mejor. En la siguiente sección se estudia cómo es posible escapar de regiones donde se tengan óptimos locales con los operadores genéticos. Por otro lado, los métodos clásicos de optimización basan su estrategia en ciertas características de los problemas y su aplicabilidad se restringe a problemas que cumplen estas características. Por ejemplo, las técnicas de gradiente necesitan la derivada para moverse en la dirección de máximo descenso. Por el contrario,

los AGs no requieren de esta información auxiliar, únicamente trabajan con la función objetivo. Por lo tanto, los AGs se pueden aplicar a una gran variedad de problemas por lo que se dice que son robustos.

1.3 Operadores Genéticos.

Los AGs ejecutan una búsqueda sobre una población constante n de individuos, donde los individuos más aptos se reproducen más, fomentando la formación de nueva información y el intercambio de información en la población a través de los operadores genéticos. Un AG sencillo está compuesto por 3 operadores genéticos:

1. Reproducción.
2. Entrecruzamiento.
3. Mutación.

REPRODUCCIÓN. La reproducción introduce un sesgo en la población al aumentar el número de individuos más aptos en la población. La reproducción se puede implementar de diversas maneras, una de ellas es la **selección proporcional** o **ruleta**. En este método, cada individuo de la población ocupa un área en la ruleta proporcional a su aptitud. Supóngase que se tiene una población de 4 individuos cuya aptitud se muestra en la siguiente tabla:

No.	cuerda	aptitud	proporción (%)
1.	01101	169	14.4
2.	11000	576	49.2
3.	01000	64	5.5
4.	10011	361	30.9
Total		1170	100

El área que ocupa cada una de estas cuerdas en la ruleta se muestra en la figura 1.2. Para reproducir la población, se gira la ruleta 4 veces que es el número de individuos que forman la población. Por ejemplo, la cuerda 1 tiene una aptitud de 169 que representa el 14.4% de la aptitud total. Por lo tanto, esta cuerda ocupa el 14.4% del área total de la ruleta. Al girar la ruleta, hay una probabilidad de 0.144 de seleccionar al individuo 1. Entonces, entre mayor sea la aptitud de un individuo, se obtienen más copias de éste en la siguiente generación y por consiguiente existe una mayor presión selectiva.

Otra manera de implementar el operador de selección es con **torneo de tamaño m**. En este caso, se seleccionan m individuos al azar y se elige al más

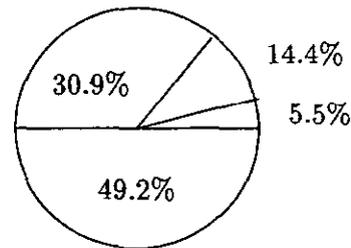


Figura 1.2: Area que ocupa cada individuo de la población en la ruleta.

apto. Este proceso se repite hasta tener una nueva población de n individuos. En este método, la presión selectiva está dada por el tamaño del torneo ya que entre mayor sea éste, más veces se selecciona al mejor individuo de la población.

ENTRECRUZAMIENTO. El entrecruzamiento entre 2 padres se realiza con una probabilidad p_c y consiste en lo siguiente. Se elige un punto k de entrecruzamiento aleatoriamente, con $1 \leq k \leq N - 1$ donde N es el tamaño de la cuerda. Los hijos están formados por la parte izquierda del punto k de un padre y por la parte derecha del punto k del otro padre. Por ejemplo, supóngase que se tienen los padres $p1$ y $p2$ con el punto de entrecruzamiento $k = 3$, los hijos $h1$ y $h2$ resultantes son:

$p1$	101 0010	$h1$	1011001
$p2$	011 1001	$h2$	0110010

El operador de entrecruzamiento tiene como función dar diversidad a la población. Si únicamente se usa selección y el mejor cromosoma no está en la población, el AG no puede encontrar el óptimo. Obsérvese que en una codificación binaria, el operador de entrecruzamiento conserva el número total de unos y ceros que hay en la población.

MUTACIÓN. La mutación es una alteración aleatoria de un alelo. Si se tiene una codificación binaria, el alelo 1 muta con una cierta probabilidad p a 0 y viceversa. La función de este operador también es dar diversidad a la población. Pero a diferencia de entrecruzamiento, las mutaciones dan seguridad en la búsqueda de soluciones aptas ya que éstas pueden obtener alelos que no se encuentran en la población o que se han perdido al aplicar algún operador genético.

Entonces, el programa de evolución de un AG es el siguiente:

```
inicio
   $t \leftarrow 0$ 
  inicializar la población al tiempo  $t$ ,  $P(t)$ 
  evaluar  $P(t)$ 
  mientras no se cumpla la condición de terminación, hacer
     $t \leftarrow t + 1$ 
    seleccionar  $P(t)$  de  $P(t - 1)$ 
    entrecruzar  $P(t)$ 
    mutar  $P(t)$ .
    evaluar  $P(t)$ 
  fin
fin.
```

Se pueden modificar algunos procedimientos en el AG para hacer la búsqueda más eficiente. Por ejemplo, en la técnica elitista la cuerda de mayor aptitud de la población, en cada iteración, pasa a la siguiente generación con probabilidad 1. De esta manera, no es posible perder esta cuerda con los operadores genéticos.

1.4 Teoría Estándar de AGs.

Una de las ventajas que tienen los AGs es que se pueden usar para resolver problemas muy diversos. Sin embargo, hay que determinar el valor de varios parámetros del AG (como la codificación, la tasa de mutación y entrecruzamiento, etc) antes de resolver el problema. Además, la eficiencia del AG en la búsqueda del óptimo depende del valor de estos parámetros y para asignarlos adecuadamente primero se debe entender cómo funciona un AG. Claramente, los AGs no obtienen la solución óptima de un problema de optimización complejo al realizar una búsqueda aleatoria sobre un espacio de soluciones. Por otro lado, tampoco se realiza una búsqueda exhaustiva al evaluar todas las combinaciones diferentes de variables o *bits* individuales. La búsqueda se realiza sobre un espacio restringido que está formado por combinaciones de *bits*. Es decir, los AGs realizan la búsqueda sobre los grados de libertad efectivos (combinaciones de *bits*) los cuales están formados por grados de libertad "microscópicos" (*bits*).

Las características de los grados de libertad efectivos dependen del paisaje de aptitud. Sin embargo, pueden tener propiedades genéricas independientes del paisaje de aptitud. La **Hipótesis de Bloques Constructores** [2], que forma parte de la teoría estándar de AGs, describe estas propiedades. Para formular esta hipótesis, es necesario introducir el concepto de **esquema**. Un esquema consiste en cuerdas parcialmente definidas. En problemas donde se usa un alfabeto binario para codificar las variables, los esquemas están definidos sobre un alfabeto ternario $\{0, 1, *\}$ donde el símbolo $*$ indica que no importa el valor que tome el *bit* en la posición correspondiente. Por lo tanto, el esquema $*01$ está aso-

ciado con las cuerdas 001 y 101. Una cuerda es un ejemplo de 2^N esquemas. Por consiguiente, un AG procesa un gran número de soluciones simultáneamente. A este fenómeno se le conoce como paralelismo implícito. Obsérvese que algunos esquemas son más específicos que otros. Por ejemplo, el esquema 011*1** contiene más información que el esquema 0*****. El número de *bits* definidos en el esquema es el **orden** N_2 del esquema. En el ejemplo anterior, el primer esquema tiene orden 4 y el último tiene orden 1. La **longitud** l de un esquema se define como la distancia entre el primer y último *bit* definido en la cuerda. Por ejemplo, el esquema 011*1** tiene longitud $l = 5$ ya que el primer *bit* definido se encuentra en la posición 1 y el último en la posición 5. La **aptitud** de un esquema se define como el promedio de las aptitudes de todas las cuerdas asociadas con el esquema bajo consideración.

En un problema que se codifica con un alfabeto binario, hay a lo más $n2^N$ esquemas en una población de tamaño n . Sin embargo, no todos los esquemas son procesados por el AG. Para saber qué esquemas se utilizan hay que considerar el efecto que tienen los operadores genéticos sobre los esquemas en cada generación. El efecto de reproducción es claro ya que las cuerdas con mayor aptitud tienen mayores probabilidades de ser seleccionadas. Entonces, en cada generación, aumentan los esquemas asociados con estas cuerdas, es decir, los esquemas con aptitud alta.

Por otro lado, entrecruzamiento no tiene ningún efecto si no se rompe el esquema durante el proceso. Considérense los esquemas 1***0 y **11*. Es más factible que se destruya el primer esquema que el segundo. Este último sólo se puede romper en la posición $k = 3$. Por lo tanto, los esquemas de longitud grande se destruyen más fácilmente con el operador de entrecruzamiento.

El operador de mutación normalmente no modifica los esquemas porque se aplica con una probabilidad p muy baja. Además, las mutaciones modifican los valores de cada *bit* y no la cuerda completa como en el caso de entrecruzamiento.

Entonces, la Hipótesis de Bloques Constructores establece que los esquemas de longitud pequeña, de orden bajo y aptitud alta se recombinan para formar cuerdas con aptitud mayor. A estos esquemas se les conocen como bloques constructores ya que un AG construye la solución óptima poco a poco al combinar estos esquemas. Estos resultados se formalizan en el **teorema de esquemas o teorema fundamental de AGs** que proporciona una cota inferior de la probabilidad de obtener un esquema ξ de tamaño N y longitud l en la generación $t + 1$:

$$P(\xi, t + 1) \geq P'(\xi, t) \left[1 - p_c \left(\frac{l-1}{N-1} \right) \right]$$

donde $P'(\xi, t)$ es la probabilidad de seleccionar el esquema ξ y p_c es la probabilidad de entrecruzamiento [2]. En el capítulo 4 se deriva una ecuación exacta

para $P(\xi, t + 1)$ y se muestra que los esquemas de longitud grande tienen preferencia sobre los de longitud pequeña en paisajes de aptitud neutrales. Es decir, se mostrará que la Hipótesis de Bloques Constructores no es completamente correcta.

Una vez que se determinen las características de los grados de libertad efectivos de un AG, se estudiará el comportamiento de éstos cuando se modifican los parámetros del algoritmo. Es decir, a través de las transiciones de fase (que se definen en el siguiente capítulo) se comprenderá mejor el papel de los operadores genéticos, como selección y mutación, en la búsqueda de soluciones aptas.

CAPÍTULO 2

Transiciones de Fase.

En este capítulo, se introduce el concepto de transición de fase el cual es muy conocido en la física. Las transiciones de fase proporcionan un marco formal a partir del cual se pueden entender los efectos que tienen los operadores genéticos en el *AG*. Para ello, se desarrolla un modelo matemático que proporciona la distribución de probabilidad para cuerdas como función del tiempo en el límite de una población infinita. Con el fin de ejemplificar el concepto de transición de fase en *AGs* se estudia el modelo de Eigen y Schuster que es un caso particular del modelo anterior. Este modelo se estudia desde el punto de vista de la física la cual proporciona un marco más formal. Además, de esta manera se puede aprovechar la herramienta de la física para comprender mejor las transiciones de fase en los *AGs*.

2.1 Transiciones de Fase en la Física.

Las transiciones de fase se pueden definir como el cambio en el comportamiento cualitativo de los grados de libertad efectivos de un sistema que se dan cuando se modifican los parámetros del sistema. Es decir, los sistemas pueden presentar un comportamiento ordenado o uno desordenado dependiendo de los parámetros. Por ejemplo, a medida que se varía la temperatura (T) y la presión (P), el agua puede existir como sólido, líquido o gas. En este caso, la fase sólida presenta más orden que la fase líquida y ésta a su vez presenta más orden que la gaseosa. Esto se muestra en la figura 2.1 la cual se conoce como **diagrama de fase**. Obsérvese que fronteras de fase bien definidas separan las regiones en que se encuentran los diferentes estados. Por ejemplo, si se cruza la curva de vapor de presión hay un brinco en la densidad del fluido de tal manera que se produce un cambio repentino del estado gaseoso al líquido. Cuando se da este comportamiento se dice que la transición de fase es de **primer orden**. Más adelante se verá que las transiciones de fase del **segundo orden** son aquéllas en las que la transición de una fase a otra se da de manera gradual. Sobre la curva de vapor de presión

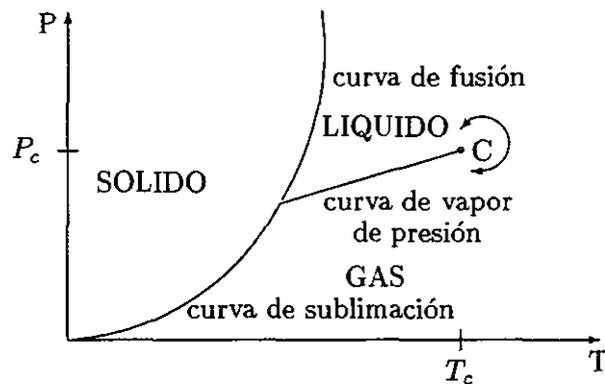


Figura 2.1: Diagrama de Fase de un Fluido.

coexisten el gas y el líquido. Sin embargo, más allá del punto crítico C se tiene un comportamiento diferente ya que es posible pasar continuamente de líquido a gas.

Para medir qué tan ordenados están los estados de un sistema se utiliza el **parámetro de orden** que se define de manera diferente dependiendo del sistema. Por ejemplo, considérese la curva de vapor de presión donde coexisten el gas y el líquido. A medida que se incrementa la temperatura a lo largo de esta curva, la diferencia entre las densidades del líquido y del gas ($\rho_l(T) - \rho_g(T)$) decrece como se ve en la figura 2.2. El parámetro de orden para este ejemplo se define como la diferencia de densidades. Cuando el parámetro de orden alcanza su máximo valor se tiene un sistema en el cual los estados están totalmente ordenados. Es decir, el sistema está compuesto únicamente por líquido o por gas. Pero a medida que la temperatura aumenta, el parámetro de orden disminuye hasta llegar a cero en T_c . En este punto, el sistema está desordenado, es decir, se puede encontrar tanto líquido como gas. Por lo tanto, la temperatura crítica T_c divide el comportamiento ordenado del desordenado en un sistema.

Otro fenómeno físico donde se da una transición de fase de primer orden es en el ferromagnetismo que consiste en lo siguiente. En algunos metales como el hierro, es posible polarizar los *spins* de los átomos (o momentos magnéticos

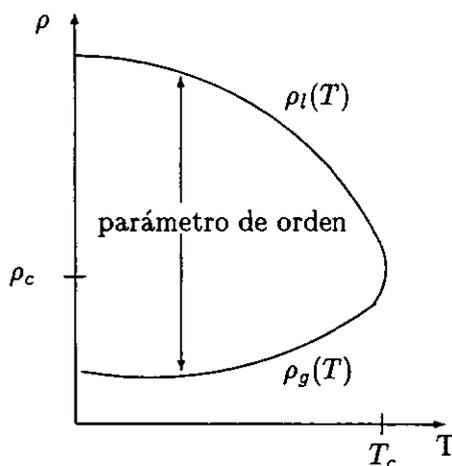


Figura 2.2: Parámetro de Orden de un Fluido.

atómicos) en la misma dirección al generar un campo magnético H . Si $H > 0$ ($H < 0$), los *spins* se orientan hacia arriba (abajo). Sin embargo, si $H = 0$ los *spins* se orientan espontáneamente en la misma dirección cuando la temperatura está por debajo de una temperatura crítica T_c (ver figura 2.3). Por arriba de esta temperatura, los *spins* se orientan aleatoriamente. Entonces, el punto crítico separa los 2 comportamientos: el primero donde los estados están ordenados (los *spins* se orientan principalmente en una dirección) y el segundo donde los estados están desordenados (los *spins* se orientan en ambas direcciones). El parámetro de orden, que en este caso es la magnetización M , indica qué tan ordenados están los *spins* a medida que se incrementa la temperatura (ver figura 2.4).

Para calcular el punto crítico T_c es necesario definir el sistema. Por ejemplo, cualitativamente el diagrama de fase para los fluidos es igual, pero la temperatura crítica T_c varía dependiendo del fluido. Sin embargo, se puede definir una nueva variable que explique el comportamiento crítico de un conjunto de sistemas que pertenecen a una **clase universal**. En el ejemplo anterior, los fluidos son la clase de interés y el oxígeno, el nitrógeno, etc. son sistemas que pertenecen a la clase de los fluidos. Si se define $t = (T - T_c)/T_c$, entonces el parámetro de orden para los fluidos es aproximadamente $(-t)^\beta$, es decir $(\rho_l - \rho_g) \sim (-t)^\beta$. La gráfica de Guggenheim (figura 2.5) presenta las curvas de coexistencia de

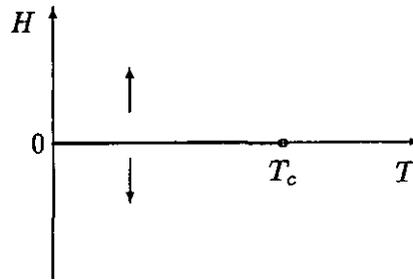


Figura 2.3: Diagrama de Fase del Ferromagnetismo.

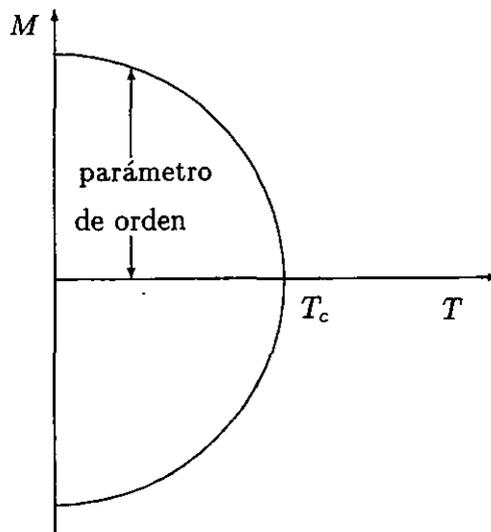


Figura 2.4: Parámetro de Orden del Ferromagnetismo.

8 fluidos diferentes pero con T/T_c y ρ/ρ_c [4]. Obsérvese que todos los datos están sobre la misma curva (en especial cerca del punto crítico) a pesar de que se tienen fluidos muy diferentes. Por lo tanto, todos estos fluidos se pueden describir con el mismo parámetro que en este caso es $\beta \approx 1/3$. Esto es de gran utilidad ya que permite hacer una clasificación de los sistemas dependiendo de su comportamiento.

Cuando se hace una simplificación del modelo de un sistema, los cálculos se facilitan. Por otro lado, se tiene una descripción general del sistema en términos de los grados de libertad efectivos en lugar de una descripción "microscópica" que en ocasiones no es adecuada. En el caso de los AGs, es más útil describir el sistema en términos de esquemas asociados con la cuerda óptima en lugar de usar las cuerdas como tales ya que es posible que algunos *bits* de la cuerda no modifiquen significativamente su aptitud. Un modelo se puede simplificar de varias maneras. Por ejemplo, las variables irrelevantes del sistema se pueden despreciar. También se puede trabajar bajo el supuesto de que estas variables tienen el mismo comportamiento. Obsérvese que al simplificar un modelo se pierde información del sistema que se puede obtener directamente de las nuevas variables "macroscópicas". En ocasiones es posible recuperar esta información al hacer un mapeo de las variables "microscópicas" a las nuevas variables. Por ejemplo, en el caso de los AGs, si se consideran todos los esquemas se obtienen las cuerdas que representan las variables originales del sistema.

La mecánica estadística establece un marco formal para el estudio de las transiciones de fase [4]. Las propiedades termodinámicas de un sistema están determinadas por la función de partición:

$$Z(T) = \sum_r e^{-\beta E_r} \quad (2.1)$$

Esta ecuación establece la probabilidad de obtener una configuración determinada de un sistema donde la suma es sobre todos los estados del sistema r con energía E_r y $\beta = \frac{1}{kT}$ con k la constante de Boltzmann.

Para ejemplificar estos conceptos se modela el fenómeno del ferromagnetismo con el modelo de Ising. Este modelo consiste de una malla donde cada punto i representa una variable S_i conocida como *spin* que puede tomar 2 valores diferentes: $S_i = \pm 1$. Los *spins* interactúan de acuerdo con la siguiente función que se conoce como Hamiltoniano:

$$\mathcal{H} = -J \sum_{i=0}^{N-1} S_i S_{i+1} - H \sum_{i=0}^{N-1} S_i \quad \text{donde } S_N = S_0$$

El primer término es responsable de que existan interacciones entre los *spins*. Si J es positivo entonces favorece que los *spins* estén alineados en la misma

dirección. De lo contrario, hay una tendencia de los *spins* S_i y S_{i+1} para orientarse en dirección opuesta. Obsérvese que la suma es sobre todos los vecinos más cercanos. En el segundo término de la ecuación no hay interacciones entre las variables. La única influencia que hay para ordenar los *spins* es el campo magnético H . Entonces, la ecuación 2.1 para este caso es:

$$Z = \sum_{\{s\}} e^{\beta J(S_0 S_1 + S_1 S_2 + \dots + S_{N-1} S_0) + \beta H(S_0 + S_1 + \dots + S_{N-1})} \quad (2.2)$$

donde $\{s\}$ representa la suma sobre todos los posibles estados del sistema, es decir, sobre $S_i = \pm 1$ para todos los *spins* S_i . A continuación, se simplifica esta ecuación como un producto de matrices. De 2.2 se tiene que

$$\begin{aligned} Z &= \sum_{\{s\}} e^{\beta J S_0 S_1 + \beta H(S_0 + S_1)/2} e^{\beta J S_1 S_2 + \beta H(S_1 + S_2)/2} \dots e^{\beta J S_{N-1} S_0 + \beta H(S_{N-1} + S_0)/2} \\ &= \sum_{\{s\}} T_{0,1} T_{1,2} \dots T_{N-1,0} \end{aligned}$$

donde

$$T_{i,i+1} = e^{\beta J S_i S_{i+1} + \beta H(S_i + S_{i+1})/2}$$

son los elementos de la siguiente matriz T :

$$T = \begin{pmatrix} e^{\beta(J+H)} & e^{-\beta J} \\ e^{-\beta J} & e^{\beta(J-H)} \end{pmatrix}$$

El primero y el segundo renglón corresponden a $S_i = 1$ y $S_i = -1$ respectivamente. Análogamente, la primera y la segunda columna corresponden a $S_{i+1} = 1$ y $S_{i+1} = -1$ respectivamente. Por ejemplo, si se tienen $N = 2$ variables de Ising entonces

$$\begin{aligned} Z &= \sum_{\substack{S_0 = \pm 1 \\ S_1 = \pm 1}} e^{\beta J S_0 S_1 + \beta H(S_0 + S_1)/2} e^{\beta J S_1 S_0 + \beta H(S_1 + S_0)/2} \\ \Rightarrow Z &= \sum_{\substack{S_0 = \pm 1 \\ S_1 = \pm 1}} T_{0,1} T_{1,0} \\ \Rightarrow Z &= e^{\beta(J+H)} e^{\beta(J+H)} + e^{-\beta J} e^{-\beta J} + e^{-\beta J} e^{-\beta J} + e^{\beta(J-H)} e^{\beta(J-H)} \end{aligned}$$

Obsérvese que esta expresión es la traza de T^2 que se denota como

$$Z_2 = \sum_{S_0=\pm 1} (T^2)_{0,0}$$

Si se desea agregar una tercera variable de Ising S_2 , entonces Z se obtiene con la traza de T^3 :

$$Z_3 = \sum_{S_0=\pm 1} (T^3)_{0,0}$$

De esta manera, la matriz T construye la malla de *spins* paso a paso por lo que se le conoce como **matriz de transferencia**. En general,

$$Z_N = \sum_{S_0=\pm 1} (T^N)_{0,0}$$

que se puede expresar también con los eigenvalores λ_i de T como

$$Z_N = \sum_i \lambda_i^N$$

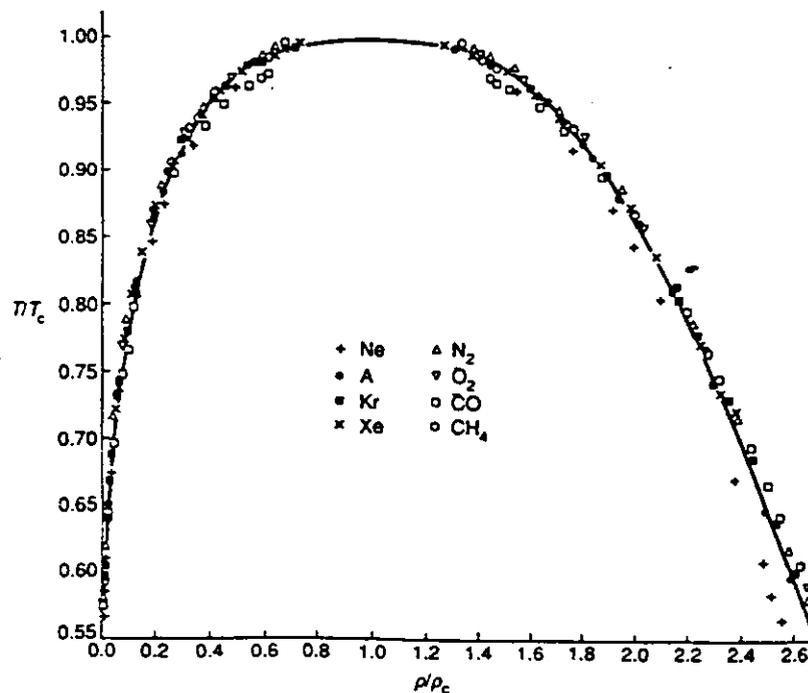


Figura 2.5: Gráfica de Guggenheim.

2.2 Modelo Matemático de un AG.

En esta sección, se desarrolla una ecuación de evolución para el valor esperado de $n(c_i, t)$, el número de cuerdas c_i en el tiempo t , en un AG con operadores de selección, mutación y entrecruzamiento. En el límite de una población infinita, esta ecuación proporciona la distribución de probabilidad para cuerdas como función del tiempo. Se empezará por desarrollar la ecuación de evolución para modelos sencillos donde las cuerdas tienen tamaño $N = 1$ y 2 . Después, se generaliza esta ecuación para cuerdas de cualquier tamaño N . Además, se analiza el papel que tienen los operadores de selección y mutación en el AG. También se estudia el proceso que sigue un AG para encontrar una solución apta con estos 2 operadores y la importancia de las transiciones de fase en este proceso.

2.2.1 Ecuaciones de Evolución para Modelos Sencillos.

Primero, se desarrollará la ecuación de evolución para el caso más sencillo donde las cuerdas tienen tamaño $N = 1$ y pueden tomar 2 valores diferentes. Es decir, sólo hay 2 estados: 1 y 0. En el lenguaje de la genética, esto es equivalente a un modelo donde los cromosomas están formados por un sólo gen y éste tiene 2 alelos diferentes.

Supóngase que el operador de selección se implementa con ruleta. Sean f_0 y f_1 las aptitudes de las cuerdas c_0 y c_1 respectivamente donde c_0 (c_1) es la cuerda cuyo único bit vale 0 (1). Para una población suficientemente grande, $P(c_i, t) = \frac{n(c_i, t)}{n}$ ($i = 0, 1$) es la probabilidad de seleccionar la cuerda c_i al tiempo t donde $n(c_i, t)$ es el número promedio de cuerdas c_i que hay en la población al tiempo t y n es el tamaño de la población. Entonces, el área que ocupa la cuerda c_i en la ruleta es $f_i P(c_i, t)$. Por lo tanto, el área total de ruleta es $\sum_i f_i P(c_i, t) = f_0 P(c_0, t) + f_1 P(c_1, t)$. Obsérvese que

$$\sum_i f_i P(c_i, t) = \sum_i \frac{f_i n(c_i, t)}{n} = \bar{f}(t)$$
 donde $\bar{f}(t)$ es la aptitud promedio de la población al tiempo t . Entonces, la probabilidad de seleccionar la cuerda c_i al tiempo $t + 1$ es:

$$P(c_i, t + 1) = \frac{f_i P(c_i, t)}{\bar{f}(t)} \quad \text{con } i = 0, 1$$

Por lo tanto, la probabilidad de seleccionar una cuerda c_i depende del número de cuerdas c_i que hay en la población así como de su aptitud.

Después de aplicar el operador de selección, se aplica el de mutación. El número esperado de cuerdas c_1 en la generación $t + 1$ depende de la proporción

de cuerdas c_1 seleccionadas en la generación t que no mutaron y de la proporción de cuerdas c_0 seleccionadas en la generación t que mutaron a c_1 . Entonces, si p es la tasa de mutación, se tiene que:

$$P(1, t+1) = P'(1, t)(1-p) + P'(0, t)p$$

donde

$$P'(c_i, t) = \frac{f_i P(c_i, t)}{\bar{f}(t)} \quad \text{con } i = 0, 1.$$

De forma análoga se obtiene $P(0, t+1)$.

Dado que las cuerdas de un *bit* no se pueden romper, el operador de entrecruzamiento no tiene ningún efecto sobre estas cuerdas.

Ahora, considérese un modelo donde las cuerdas son de tamaño $N = 2$ y donde cada *bit* puede tomar $K = 2$ valores diferentes. El número total de estados para este modelo es $K^N = 2^2 = 4$, a saber 00, 01, 10 y 11.

La probabilidad de seleccionar una cuerda c_i es independiente del tamaño de ésta, por lo tanto

$$P(11, t+1) = P'(11, t) = \frac{f_{11} P(11, t)}{\bar{f}(t)}$$

El procedimiento para obtener el número esperado de cuerdas 11 después de aplicar los operadores de selección y mutación es igual al caso donde $N = 1$. Sin embargo, la probabilidad de mutación $p(k)$ se aplica sobre el *bit* k ($k = 1, 2$) de la cuerda. Si se considera una probabilidad de mutación constante $p(k) = p$ con $k = 1, 2$ entonces la probabilidad de que la cuerda 11 no tenga mutaciones es $(1-p)^2$. Por lo tanto,

$$P(11, t+1) = P'(11, t)(1-p)^2 + P'(10, t)p(1-p) + P'(01, t)p(1-p) + P'(00, t)p^2$$

Para calcular el número esperado de cuerdas 11 que se obtienen con el operador de entrecruzamiento, se deben tomar en cuenta los efectos constructivo y destructivo de este operador. Es decir, la cuerda 11 se puede construir si se aplica el operador de entrecruzamiento a las cuerdas 10 y 01:

$$\begin{array}{l} 10 \rightarrow 00 \\ 01 \quad 11 \end{array}$$

Por otro lado, la única manera de destruir la cuerda 11 es al entrecruzar esta cuerda con 00:

$$\begin{array}{ccc} 11 & \rightarrow & 10 \\ 00 & & 01 \end{array}$$

Sea p_c la probabilidad de entrecruzamiento. Entonces, la probabilidad de obtener la cuerda 11 en el tiempo $t + 1$ al aplicar selección y después entrecruzamiento es:

$$P(11, t + 1) = P'(11, t) - p_c P'(11, t) P'(00, t) + p_c P'(10, t) P'(01, t)$$

Si se considera un *AG* donde se aplica selección, después entrecruzamiento y por último mutación; la ecuación de evolución para la cuerda 11 es:

$$P(11, t + 1) = P_c(11, t)(1 - p)^2 + P_c(10, t)p(1 - p) + P_c(01, t)p(1 - p) + P_c(00, t)p^2$$

donde $P_c(ij, t)$ es la probabilidad de obtener la cuerda ij en el tiempo t al aplicar selección y entrecruzamiento. Obsérvese que para pasar de una generación t a la siguiente generación $t + 1$ se deben aplicar los 3 operadores genéticos. Si $ij = 11$ se tiene que

$$P_c(11, t + 1) = P'(11, t) - p_c P'(11, t) P'(00, t) + p_c P'(10, t) P'(01, t)$$

2.2.2 El Efecto de la Selección en un *AG*.

Se sabe que la probabilidad de seleccionar una cuerda c_i con ruleta al tiempo $t + 1$ en una población suficientemente grande es:

$$P(c_i, t + 1) = P'(c_i, t) = \frac{f_i P(c_i, t)}{\bar{f}(t)} \quad (2.3)$$

Intuitivamente, se espera que las cuerdas con una aptitud mayor que la del promedio de la población aumenten en cada generación. Es decir, se espera que

si C_i es tal que $f_i > \bar{f}(t)$ entonces C_i aumenta en cada generación. Esto determina la ventaja selectiva de la cuerda C_i que se define como:

$$\delta f_i = \frac{f_i}{\bar{f}(t)} - 1$$

Se puede calcular analíticamente la tasa de cambio del número esperado de cuerdas C_i como función del tiempo. Para ello, se obtiene la ecuación diferencial de la siguiente manera:

$$\frac{dP(C_i, t)}{dt} = \lim_{\delta t \rightarrow 0} \frac{P(C_i, t + \delta t) - P(C_i, t)}{\delta t}$$

ya que t no es continuo porque en un AG se consideran generaciones. Entonces,

$$\begin{aligned} \frac{dP(C_i, t)}{dt} &= P(C_i, t + 1) - P(C_i, t) \\ &= \left[\frac{f_i}{\bar{f}(t)} - 1 \right] P(C_i, t) \\ \Rightarrow P(C_i, t) &= P(C_i, 0) e^{\int_0^t \delta f_i dt'} \end{aligned}$$

En una población inicial aleatoria,

$$P(C_i, 0) \simeq P(C_j, 0) \quad \forall i \neq j$$

Por lo que el valor de $P(C_i, t)$ depende principalmente de δf_i . Se tienen 3 casos:

1. Si $\delta f_i > 0$, el número de cuerdas C_i aumenta en promedio en cada generación.
2. Si $\delta f_i = 0$, el número de cuerdas C_i permanece constante en promedio en cada generación.
3. Si $\delta f_i < 0$, el número de cuerdas C_i disminuye en promedio en cada generación.

Si las cuerdas con mayor ventaja selectiva tienden a aumentar en cada generación, entonces se esperaría que la aptitud promedio en la población también aumente en cada generación. Es decir,

$$\bar{f}(t+1) \geq \bar{f}(t)$$

A esta propiedad se le conoce como Teorema Fundamental de Selección Natural de Fisher y se demuestra de la siguiente manera.

DEMOSTRACIÓN. Se tiene que

$$P(c_i, t+1) = \frac{n(c_i, t+1)}{n} \quad (2.4)$$

$$\Rightarrow n(c_i, t+1) = \frac{f_i n(c_i, t)}{\bar{f}(t)} \quad (2.5)$$

Por otro lado,

$$\begin{aligned} \sum_i (f_i - \bar{f}(t))^2 P(c_i, t) &\geq 0 \\ \Rightarrow \sum_i f_i^2 P(c_i, t) - 2\bar{f}(t) \sum_i f_i P(c_i, t) \\ &\quad + (\bar{f}(t))^2 \sum_i P(c_i, t) \geq 0 \\ \Rightarrow \bar{f}^2(t) - 2(\bar{f}(t))^2 + (\bar{f}(t))^2 &\geq 0 \\ \Rightarrow \bar{f}^2(t) &\geq (\bar{f}(t))^2 \end{aligned} \quad (2.6)$$

Entonces,

$$\begin{aligned} \bar{f}(t+1) &= \frac{\sum_i f_i n(c_i, t+1)}{n} \\ &= \frac{\sum_i f_i \left[\frac{n(c_i, t) f_i}{\bar{f}(t)} \right]}{n} \quad (\text{por 2.5}) \\ &= \frac{\sum_i f_i^2 n(c_i, t)}{n \bar{f}(t)} = \frac{\sum_i f_i^2 n(c_i, t)}{\sum_i f_i n(c_i, t)} \left(\frac{n}{n} \right) \\ \Rightarrow \bar{f}(t+1) &= \frac{\bar{f}^2(t)}{\bar{f}(t)} \geq \frac{(\bar{f}(t))^2}{\bar{f}(t)} \quad (\text{por 2.6}) \\ \Rightarrow \bar{f}(t+1) &\geq \bar{f}(t) \end{aligned}$$

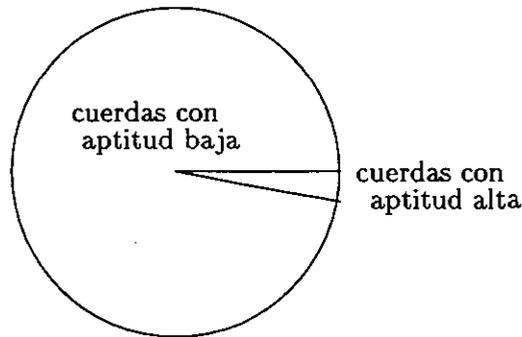


Figura 2.6: Área de las cuerdas en la ruleta en una población pequeña.

Se ha mostrado que en una población suficientemente grande, la selección aumenta el número de cuerdas más aptas en cada generación. De esta manera, la población en el punto estable (es decir, en el límite cuando la población no se modifica) estará formada únicamente por la cuerda más apta. Es decir, el efecto que tiene la selección en una población suficientemente grande es el de ordenarla de tal manera que toda la población está formada por la misma cuerda. Sin embargo, este comportamiento no necesariamente se da cuando el tamaño de la población es pequeño. En este caso, se pueden presentar muchas fluctuaciones en la aptitud promedio de la población debido al efecto de tamaño finito de la población. Si la población es pequeña (con respecto al número total de estados posibles del sistema), es probable que las cuerdas más aptas aparezcan en proporciones muy bajas en la población. Incluso, si la población inicial se genera aleatoriamente el óptimo puede no encontrarse en la población inicial en especial si el número de estados del sistema es muy grande. Entonces, la probabilidad de seleccionar las cuerdas aptas de la población disminuye si su concentración relativa dentro de la población es muy baja. Esta situación se ejemplifica en la figura 2.6 donde las cuerdas menos aptas de la población ocupan un área mayor en la ruleta.

A continuación, se estudiará cómo ordena la selección a la población en 3 paisajes de aptitud diferentes cuando la población es infinita. El primero, es un paisaje plano donde todas las cuerdas tienen la misma aptitud (ver figura 2.7). En este caso, no hay ninguna cuerda que tenga ventaja selectiva sobre las demás ya que $f_i = f \forall i$. Por lo tanto,

$$P(c_i, t + 1) = P(c_i, t)$$

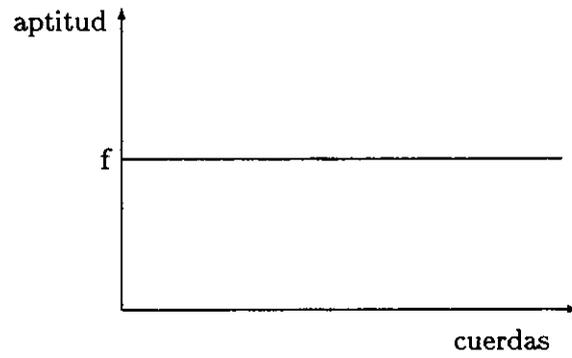


Figura 2.7: Paisaje de Aptitud Plano.

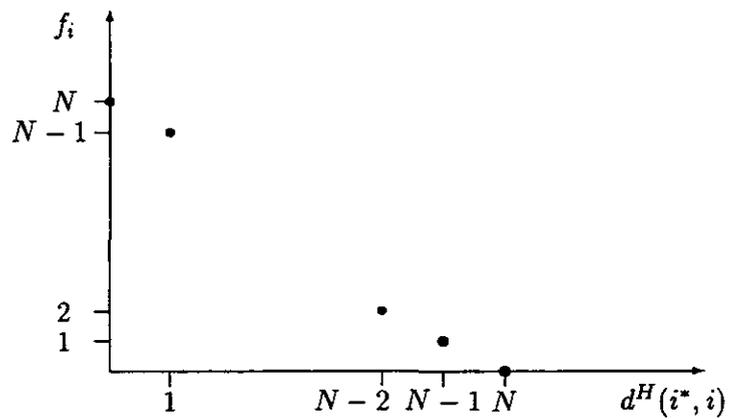


Figura 2.8: Paisaje de aptitud donde se cuenta el número de unos.

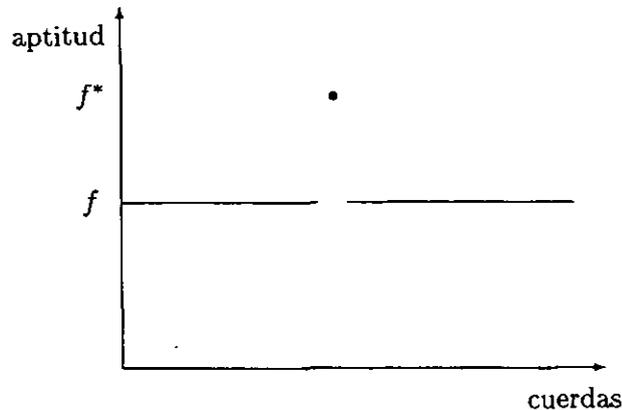


Figura 2.9: Paisaje de Aptitud Plano con un Pico.

Dado que $P(c_i, t)$ no se modifica con el tiempo, la selección no ordena la población. Entonces, el efecto de la selección depende del paisaje de aptitud.

Ahora, considérese un paisaje de aptitud donde la aptitud de cada cuerda es igual al número de unos que tiene ésta. Por ejemplo, la cuerda 00111 tiene aptitud 3. Dado que $f_i \in \{0, 1, \dots, N\}$, la cuerda (c_{i^*}) con mayor aptitud es aquella en la que todos sus *bits* valen 1. Se puede medir qué tan diferente es una cuerda c_i respecto a c_j si se cuenta el número de *bits* de c_i que tienen un valor diferente al de la cuerda c_j . A esta medida se le conoce como **distancia de Hamming** y se denota por $d^H(i, j)$. Por ejemplo, si $c_i = 10100$ y $c_j = 11011$ entonces $d^H(i, j) = 4$. Entonces, se puede representar gráficamente este paisaje de aptitud como se muestra en la figura 2.8 donde $c_{i^*} = 11 \dots 1$. Obsérvese que las cuerdas se agrupan en el eje x de acuerdo con su distancia de Hamming a la cuerda óptima. Por ejemplo, hay N cuerdas con $d^H(i^*, i) = 1$. También es posible representar este paisaje de aptitud en un hipercubo de dimensión N como se muestra en la figura 2.12 donde $N = 3$.

En este paisaje de aptitud, la selección ordena la población de tal manera que en el punto estable está formada únicamente por la cuerda $c_{i^*} = 11 \dots 1$ ya que ésta es la que tiene mayor ventaja selectiva.

El tercer ejemplo es un paisaje de aptitud plano con un pico (ver figura 2.9). Esto significa que todas las cuerdas tienen la misma aptitud f excepto la cuerda óptima que se conoce como **secuencia maestra** y tiene una aptitud $f^* > f$. Supóngase que la secuencia maestra es 0101 \dots 01. Entonces, en el punto estable la población estará formada únicamente por la cuerda 0101 \dots 01.

2.2.3 El Efecto de Selección y Mutación en un AG.

Considérense cuerdas de longitud N . Entonces, la ecuación de evolución de la cuerda c_i bajo los operadores de selección y mutación es la siguiente:

$$P(c_i, t+1) = \mathcal{P}(c_i)P'(c_i, t) + \sum_{c_j \neq c_i} \mathcal{P}(c_j \rightarrow c_i)P'(c_j, t) \quad (2.7)$$

donde $\mathcal{P}(c_i)$ es la probabilidad de que c_i no mute y $\mathcal{P}(c_j \rightarrow c_i)$ es la probabilidad de que la cuerda c_j mute a la cuerda c_i . Si $p(k)$ es la probabilidad de mutación del bit k , entonces

$$\mathcal{P}(c_i) = \prod_{k=1}^N (1 - p(k))$$

Si se supone una tasa de mutación constante $p(k) = p$, se tiene que

$$\mathcal{P}(c_i) = (1 - p)^N$$

Por otro lado, la probabilidad de que la cuerda c_j mute a la cuerda c_i es:

$$\mathcal{P}(c_j \rightarrow c_i) = \prod_{k \in \{C_j - C_i\}} p(k) \prod_{k \in \{C_j - C_i\}^c} (1 - p(k))$$

donde $\{C_j - C_i\}$ es el conjunto de bits de c_j diferentes a los de c_i y $\{C_j - C_i\}^c$ es el complemento de este conjunto, es decir, los bits que tienen en común c_i y c_j . Si $p(k) = p$ es constante, entonces

$$\mathcal{P}(c_j \rightarrow c_i) = p^{d^H(i,j)} (1 - p)^{N - d^H(i,j)}$$

A continuación, se estudiará el efecto que tienen los operadores de selección y mutación en un AG. Si la tasa de mutación es muy alta entonces existe una competencia entre selección y mutación donde la selección ordena la población y la mutación la desordena. La selección incrementa el número de individuos más aptos y la mutación produce nuevos individuos que generalmente tienen una aptitud menor. Si un gen mutado tiene un efecto muy desventajoso, entonces la selección actúa fuertemente sobre el individuo y como consecuencia su frecuencia en la población permanece baja. Si el gen es ligeramente desventajoso, el efecto que tiene la selección sobre el individuo es menor y por lo tanto aparece con mayor frecuencia en la población que una mutación más desventajosa.

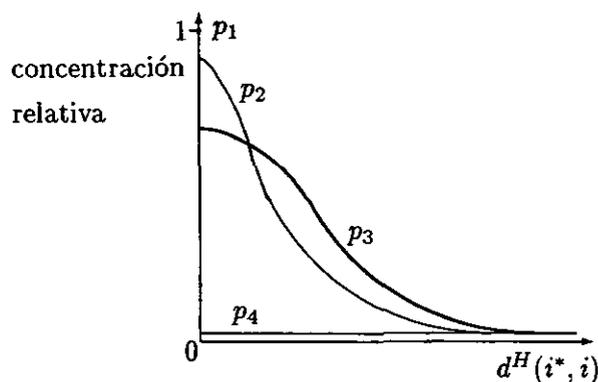


Figura 2.10: Distribución estacionaria para diferentes tasas de mutación p con $p_1 = 0 < p_2 < p_3 < p_4 = 1$. En la gráfica, se muestra la concentración relativa de las cuerdas en el punto estable agrupadas por su distancia de Hamming respecto al óptimo.

Entonces, en ausencia de mutaciones, la cuerda con la mayor tasa de reproducción aumenta hasta dominar la población de tal manera que todas las cuerdas que sobreviven son idénticas. Si hay mutaciones, se forman continuamente nuevas cuerdas que no tienen aptitud óptima. En este caso, la concentración relativa de cada cuerda dentro de la población converge a una distribución estacionaria donde las cuerdas diferentes a la cuerda óptima están presentes en cantidades significativas. A esta distribución estacionaria se le conoce como **cuasi-especie**.

La distribución estacionaria depende de la manera en que la aptitud disminuye con sucesivas mutaciones de la cuerda óptima. En general, la distribución se extiende a medida que la tasa de mutación se incrementa (ver figura 2.10). Para tasas de mutación grandes, la selección no tiene ningún efecto y la distribución estacionaria se distribuye a lo largo de todo el espacio de estados. En algunos casos, hay una transición bien definida entre una distribución localizada y una no localizada a medida que la tasa de mutación aumenta. Esta transición se conoce como **umbral de error** o **punto de transición de fase**.

Para ejemplificar el concepto de transición de fase, considérese el modelo de cromosomas constituidos por un sólo gen descrito en la sección 2.2.1. Supóngase que el gen tiene muchos alelos de tal manera que un alelo mutado no puede mutar para volver a formar al alelo óptimo. El alelo óptimo tiene aptitud $f^* = 1$

y los demás alelos mutados tienen una aptitud de $f = 1 - s$ con s constante. Sea $X(t)$ la concentración relativa del alelo óptimo al tiempo t y $Y(t) = 1 - X(t)$ la concentración relativa combinada de los demás alelos mutados. Entonces, la ecuación de evolución para el alelo óptimo es:

$$X(t+1) = \frac{X(t)(1-p)}{X(t) + (1-s)Y(t)}$$

Sea $l = \lim_{t \rightarrow \infty} X(t)$. En la distribución estacionaria se tiene que $l = X(t+1) = X(t)$, por lo tanto

$$l = \frac{l(1-p)}{l + (1-s)(1-l)}$$

$$\Rightarrow l = 1 - \frac{p}{s}$$

dado que $l \geq 0 \Rightarrow p \leq s$. Entonces

$$\lim_{t \rightarrow \infty} X(t) = \begin{cases} 1 - \frac{p}{s} & \text{si } p \leq s \\ 0 & \text{si } p > s \end{cases}$$

La gráfica 2.11 muestra la concentración relativa del alelo óptimo en la distribución estacionaria para diferentes tasas de mutación p . Se fijó s en 0.5. En la gráfica se puede observar una transición de fase entre un régimen ordenado (dominado por selección) y uno desordenado (dominado por mutación). El umbral de error se da en $p = 0.5 (= s)$. Para tasas de mutación mayores que esta tasa crítica, hay una pérdida en la información genética obtenida previamente.

La búsqueda que realiza un AG se puede ver como un balance entre exploración (de nuevas soluciones) y explotación (de soluciones aptas encontradas previamente). Cuando la tasa de mutación es alta, el proceso de búsqueda no puede explotar las soluciones aptas obtenidas con anterioridad y por lo tanto se transforma en una búsqueda aleatoria. El umbral de error es importante en un AG ya que determina un balance crítico entre exploración y explotación. Para maximizar la exploración, la tasa de mutación debe ser lo más alta posible pero no debe exceder el umbral de error. Cabe hacer notar que después de varias generaciones, cuando la aptitud promedio de la población se ha incrementado, es conveniente usar la técnica elitista. De no hacerlo, es posible que se pierdan las cuerdas óptimas encontradas previamente.

A continuación, se describe el papel que tienen las transiciones de fase en la búsqueda del óptimo. Se mostrará que la optimización evolutiva puede ser guiada a los picos (o puntos de mayor aptitud) del paisaje de aptitud a través de mutaciones sesgadas cuando la población es suficientemente grande. Las mutaciones son aleatorias, sin embargo, hay un sesgo en la distribución de mutaciones

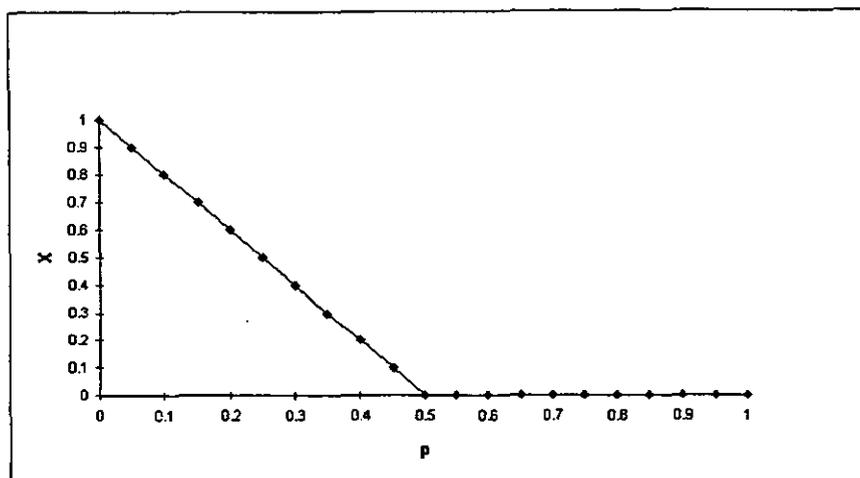


Figura 2.11: Transición de Fase para el Modelo de un Gen.

producido por la aptitud de éstas, de tal manera que las mutaciones ventajosas aparecen en proporciones mayores que las desventajosas.

Generalmente, la tasa de mutación es pequeña. Como consecuencia, el sistema se ordena relativamente rápido alrededor de un óptimo local. Es decir, se genera una distribución cuasi-especie. Por el contrario, la generación de una nueva mutación con ventaja selectiva es un proceso lento. Estas mutaciones se originan a partir de mutaciones en la periferia de la distribución cuasi-especie. Es decir, a una distancia de Hamming relativamente grande con respecto a la cuerda ventajosa.

La concentración relativa de una determinada mutación depende de la aptitud del óptimo local así como de la aptitud de las demás mutaciones en la distribución cuasi-especie. Si hay 2 mutaciones con la misma distancia de Hamming respecto al óptimo local, la mutación de mayor aptitud tendrá una concentración relativa en la población más alta que la de menor aptitud. Entonces, es más probable que el AG elija las rutas de optimización cuyas cuerdas son más aptas. Para ilustrar este proceso, se ordenan las mutaciones de acuerdo a la distancia de Hamming que hay entre éstas en un espacio N-dimensional como se muestra en la figura 2.12. En los caminos evolutivos, se puede brincar de una cuerda a otra sólo si la distancia de Hamming entre estas 2 cuerdas es 1. La aptitud de cada cuerda se representa con f_i donde $f_i < f_j \forall i < j$. La cuerda óptima es 011. Para ir de la cuerda 101 a la óptima hay varios caminos evolutivos. Sin embargo, es más factible elegir el camino 101 - 001 - 011 ya que en cada paso la aptitud de la cuerda aumenta. Entonces, el proceso evolutivo no es una búsqueda aleatoria de las mutaciones ventajosas. Si fuera así, habría una probabilidad muy baja

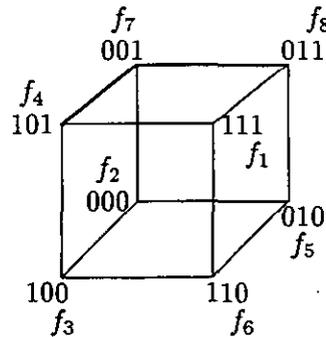


Figura 2.12: Representación de los estados de un sistema respecto a sus mutaciones.

de encontrar el óptimo global cuando el número de estados del sistema es muy grande. Como consecuencia, en el proceso evolutivo no importa únicamente el óptimo local sino toda la distribución cuasi-especie.

La optimización evolutiva se puede ver como una serie de estabilizaciones y desestabilizaciones de la distribución cuasi-especie (es decir, como una serie de transiciones de fase) que, en un ambiente constante, se asocian con un incremento en la aptitud. El proceso es guiado por una distribución de mutaciones en cada cuasi-especie. Las mutaciones en la periferia de la distribución cuasi-especie tienen una aptitud menor que el óptimo local correspondiente. Sin embargo, las mutaciones menos desventajosas generalmente se encuentran en un área de aptitud alta. Por lo tanto, evitan el área con aptitud baja y así la optimización procede a lo largo de un número de caminos limitados. El umbral de error es de gran importancia ya que el progreso evolutivo es mayor cerca de éste. Si se sobrepasa ocasionalmente este límite, se puede escapar de determinadas distribuciones y encontrar picos más altos en el paisaje de aptitud. Esto es similar al proceso que se sigue en el recocido simulado para realizar una búsqueda global.

El fenómeno de transición de fase no se da siempre. A continuación se analizan varios paisajes de aptitud estáticos. Por ejemplo, en el paisaje de aptitud plano no hay transición de fase porque todas las cuerdas tienen la misma aptitud. En el modelo donde la aptitud de una cuerda es igual al número de unos de ésta, tampoco hay transición de fase por la siguiente razón. Si se modifica el valor de un *bit* de una cuerda, no se produce un cambio muy grande en la aptitud de ésta. Por lo tanto, el efecto que domina en este caso es el de selección dado que las mutaciones no son muy destructivas.

En el modelo anterior, cada *bit* tiene una contribución en la aptitud de la

cuerda que es independiente de los otros *bits*. Es decir, si un *bit* tiene un valor de uno entonces contribuye en una unidad y si tiene un valor de cero, no contribuye en la aptitud de la cuerda. Pero esta contribución es independiente del valor que tengan los otros *bits*. Sin embargo, esta es una idealización. En general, si se tiene una cuerda con N *bits*, la contribución en la aptitud de un *bit* depende de los valores de algunos de los $N - 1$ *bits* restantes. A esta dependencia se le conoce como **interacciones epistáticas**. En el paisaje de aptitud plano con un pico, la contribución del valor de un *bit* en la aptitud de la cuerda depende de los valores de los $N - 1$ *bits* restantes. En este caso, la secuencia maestra es $11 \cdots 1$ y por lo tanto todos los *bits* deben valer 1 para tener la cuerda óptima de aptitud f^* . De otra manera, la aptitud de la cuerda es $f < f^*$. Es decir, no hay diferencia entre la cuerda $011 \cdots 1$ y la cuerda $00 \cdots 0$ para construir la secuencia maestra. Como consecuencia, un cambio en el valor de un *bit* de la secuencia maestra produce un cambio muy grande en la aptitud de la cuerda de tal manera que se pasa de una cuerda de aptitud máxima a una cuerda de aptitud mínima. Entonces, el efecto destructivo de la mutación se amplifica. Si la tasa de mutación es alta, la selección no tiene ningún efecto. Por lo tanto, en este paisaje de aptitud sí se tiene una transición de fase.

2.2.4 Ecuación de Evolución con Todos los Operadores Genéticos.

En esta sección se incorpora el operador de entrecruzamiento a la ecuación de evolución para cuerdas de tamaño N . Los efectos de este operador se estudiarán con más detalle en el capítulo 4.

La ecuación de evolución considerando únicamente los efectos de reproducción y entrecruzamiento es:

$$\begin{aligned}
 P(c_i, t+1) &= P'(c_i, t) \\
 &- \frac{p_c}{N-1} \sum_{c_j \neq c_i} \sum_{k=1}^{N-1} C_{c_i, c_j}^{(1)}(k) P'(c_i, t) P'(c_j, t) \\
 &+ \frac{p_c}{N-1} \sum_{c_j \neq c_i} \sum_{c_l \neq c_i} \sum_{k=1}^{N-1} C_{c_j, c_l}^{(2)}(k) P'(c_j, t) P'(c_l, t) \quad (2.8)
 \end{aligned}$$

El segundo término de esta ecuación representa el término destructivo del operador de entrecruzamiento y el tercer término representa el término constructivo. La cuerda c_i se destruye con el operador de entrecruzamiento si se cruza con una cuerda c_j tal que

$$d_L^H(i, j) > 0 \quad \text{y} \quad d_R^H(i, j) > 0$$

donde $d_L^H(i, j)$ ($d_R^H(i, j)$) es la distancia de Hamming entre las cuerdas C_i y C_j tomando en cuenta sólo la parte izquierda (derecha) del punto de entrecruzamiento k . Entonces, $C_{C_i, C_j}^{(1)}(k)$ se define como:

$$C_{C_i, C_j}^{(1)}(k) = \theta(d_L^H(i, j))\theta(d_R^H(i, j))$$

donde

$$\theta(x) = \begin{cases} 1 & \text{si } x > 0 \\ 0 & \text{si } x = 0 \end{cases}$$

Obsérvese que si C_i se destruye, $C_{C_i, C_j}^{(1)}(k) = 1$. Para construir la cuerda C_i a partir de las cuerdas C_j y C_l se debe cumplir la siguiente condición:

$$(d_L^H(i, j) = 0 \quad \text{y} \quad d_R^H(i, l) = 0) \quad \text{ó} \quad (d_R^H(i, j) = 0 \quad \text{y} \quad d_L^H(i, l) = 0)$$

Entonces, el término constructivo $C_{C_j, C_l}^{(2)}(k)$ se define como

$$C_{C_j, C_l}^{(2)}(k) = \frac{1}{2} \left[\delta(d_L^H(i, j))\delta(d_R^H(i, l)) + \delta(d_R^H(i, j))\delta(d_L^H(i, l)) \right]$$

donde

$$\delta(x) = \begin{cases} 1 & \text{si } x = 0 \\ 0 & \text{si } x \neq 0 \end{cases}$$

Por lo tanto, si C_i se puede construir a partir de C_j y C_l entonces $C_{C_j, C_l}^{(2)}(k) \neq 0$.

De la ecuación 2.8 se observa que el operador de entrecruzamiento se aplica con una probabilidad p_c y la probabilidad de elegir un determinado punto k ($1 \leq k \leq N-1$) de entrecruzamiento es $\frac{1}{N-1}$.

Entonces, la probabilidad de obtener una cuerda C_i en la generación $t+1$ en un AG donde después de seleccionar la población se aplica entrecruzamiento y por último mutación es:

$$P(C_i, t+1) = \mathcal{P}(C_i)P_c(C_i, t) + \sum_{C_j \neq C_i} \mathcal{P}(C_j \rightarrow C_i)P_c(C_j, t) \quad (2.9)$$

donde

$$\begin{aligned}
 P_c(c_i, t) &= P'(c_i, t) \\
 &- \frac{p_c}{N-1} \sum_{C_j \neq C_i} \sum_{k=1}^{N-1} C_{c_i c_j}^{(1)}(k) P'(c_i, t) P'(c_j, t) \\
 &+ \frac{p_c}{N-1} \sum_{C_j \neq C_i} \sum_{C_l \neq C_i} \sum_{k=1}^{N-1} C_{c_j c_l}^{(2)}(k) P'(c_j, t) P'(c_l, t)
 \end{aligned}$$

2.3 El Modelo de Eigen y Schuster.

En esta sección, se estudiará el modelo de Eigen y Schuster para la evolución de moléculas que se autorepican [13]. En este modelo, las moléculas se representan con cuerdas (s_1, s_2, \dots, s_N) de longitud N fija. Las variables s_i representan los diferentes monómeros que forman las moléculas. Si las moléculas son de ADN entonces s_i representa las 4 bases (guanina, adenina, citocina y timina) que forman el ADN. Sin embargo, sólo se consideran variables binarias (± 1) que representan las purinas (guanina y adenina) y las pirimidinas (citocina y timina). Se denota por S_k , con $k = 1, 2, \dots, 2^N$, a las diferentes cuerdas y por X_k a la concentración de la molécula S_k en la población. Además, se considera una población grande de moléculas. Este modelo se puede aplicar a sistemas más complejos como los virus donde los individuos se auto-reproducen. Sin embargo, este modelo no considera entrecruzamiento.

La evolución en el tiempo de la concentración relativa de estas moléculas es:

$$\frac{dX_k}{dt} = \sum_{j \neq k} W_{kj} X_j + [W_{kk} - D_k] X_k = \sum_j W'_{kj} X_j$$

donde W_{kj} son los elementos de la matriz de reproducción W que indican la probabilidad de obtener la molécula S_k como resultado de la reproducción de S_j . Esta matriz es diagonal si cada molécula produce una o varias copias perfectas de sí misma. Sin embargo, el modelo introduce la posibilidad de mutación en el proceso de reproducción. Por lo que los elementos de W que no están en la diagonal tienen valores positivos pequeños. Por otro lado, D_k representa la probabilidad de muerte de cada molécula.

Leuthäusser [6] realizó un mapeo de la evolución de las moléculas en un sistema termodinámico en equilibrio. La termodinámica estudia las propiedades de la materia cuando son afectadas por un cambio de temperatura entre otros factores. Para realizar este mapeo, se hace discreto el tiempo, es decir se consideran generaciones, por lo que en lugar de tener $X_k(t)$ se tiene $X_k(i)$ con $i = 0, 1, \dots$

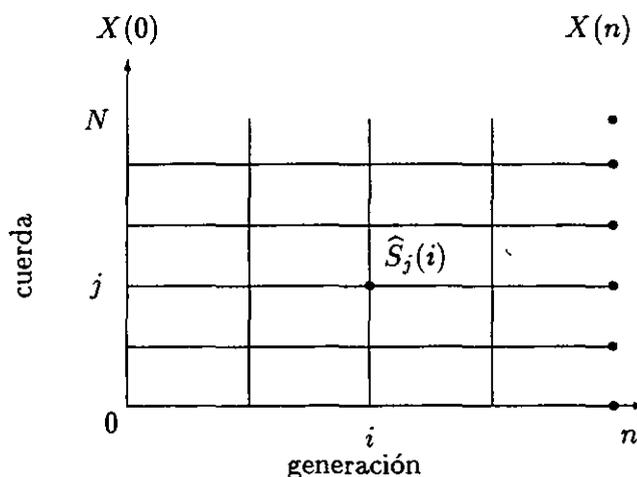


Figura 2.13: Malla donde se representa la evolución de las moléculas.

Esto representa la concentración relativa de la molécula S_k en la generación i al empezar con una población $X_k(0)$. La matriz W_{kj} representa la probabilidad para cada molécula S_j de producir la molécula S_k en la siguiente generación. Entonces, la población después de n generaciones está dada por

$$X(n) = W^n X(0) \quad (2.10)$$

donde $X(i) = (X_1(i), X_2(i), \dots, X_{2N}(i))$. La matriz W se puede ver como una matriz de transferencia que proporciona la probabilidad de obtener $X(i+1)$ en términos de su vecino más cercano en tiempo $X(i)$ ya que $X(i+1) = WX(i)$. El eigenvalor más grande de W proporciona la tasa de crecimiento de la molécula más numerosa en el estado estacionario, es decir, cuando la población de moléculas casi no se modifica. El eigenvector correspondiente indica la composición de la población en términos de la concentración relativa de cada molécula [6].

La distribución de las moléculas en la n -ésima generación depende de todas las posibles trayectorias entre el estado inicial $X(0)$ y el estado final $X(n)$. La ecuación 2.10 considera implícitamente todos los caminos evolutivos. Cada camino evolutivo se puede describir en una malla de 2 dimensiones como se muestra en la figura 2.13. Cada punto en la red $\hat{S}_j(i)$ representa la concentración relativa de un cierto valor en la j -ésima posición de todas las moléculas de la población en la generación i . Por ejemplo para $N = 4$, si se tiene la población $(1, 1, -1, 1)$, $(1, 1, 1, 1)$, $(-1, 1, -1, 1)$ y $(1, 1, -1, -1)$ en la i -ésima generación,

entonces la concentración relativa de 1's que hay en la posición 3 en toda la población es $\hat{S}_3(i) = -\frac{1}{2}$. Las propiedades importantes del sistema se obtienen de la distribución de la última columna ($i = n$) o **superficie**.

En este caso, la función de partición está dada por:

$$W_{jk} = \exp(-\beta h[S_j, S_k]) \quad (2.11)$$

donde $H = \sum_{i=0}^{n-1} h[S(i), S(i+1)]$ es el Hamiltoniano. A continuación se encontrará el valor de los parámetros $\beta h[S_j, S_k]$ de la termodinámica en términos del modelo de Eigen y Schuster. Dado que en este modelo sólo se considera reproducción y mutación, W_{jk} se puede obtener a partir de la ecuación 2.7

$$W_{jk} = A[S_j](1-p)^N \left(\frac{p}{1-p}\right)^{d^H(S_j, S_k)} \quad (2.12)$$

donde $d^H(S, S') = \frac{1}{2}[N - \sum_{k=1}^N S_k S'_k]$ para moléculas S y S' y $A[S]$ es la aptitud o tasa de reproducción de la molécula S . De 2.11 y 2.12 se tiene que

$$\begin{aligned} \exp(-\beta h[S_j(i), S_j(i+1)]) &= A[S_j(i)](1-p)^N \left(\frac{p}{1-p}\right)^{d^H(S_j(i), S_j(i+1))} \\ \Rightarrow -\beta h[S_j(i), S_j(i+1)] &= \ln \left\{ A[S_j(i)](1-p)^N \left(\frac{p}{1-p}\right)^{d^H(S_j(i), S_j(i+1))} \right\} \\ &= \ln A[S_j(i)] + \ln(1-p)^N \\ &\quad + d^H(S_j(i), S_j(i+1)) \ln \left(\frac{p}{1-p}\right) \\ &= \ln A[S_j(i)] + \ln(1-p)^N \\ &\quad + \frac{1}{2} \left[N - \sum_{j=1}^N S_j(i) S_j(i+1) \right] \ln \left(\frac{p}{1-p}\right) \\ &= \ln A[S_j(i)] + \ln(1-p)^N \\ &\quad + \frac{N}{2} \ln \left(\frac{p}{1-p}\right) - \frac{1}{2} \sum_{j=1}^N S_j(i) S_j(i+1) \ln \left(\frac{p}{1-p}\right) \\ &= \ln A[S_j(i)] + \frac{N}{2} \ln [(1-p)p] \\ &\quad + \sum_{j=1}^N S_j(i) S_j(i+1) \ln \sqrt{\frac{(1-p)}{p}} \end{aligned}$$

El parámetro β se define como $\beta = \ln \sqrt{\frac{(1-p)}{p}}$ [5], entonces

$$\begin{aligned}
-\beta h[S_j(i), S_j(i+1)] &= \beta \sum_{j=1}^N S_j(i) S_j(i+1) + \ln A[S_j(i)] + \frac{N}{2} \ln [(1-p)p] \\
\Rightarrow -\beta H &= \sum_{i=0}^{n-1} \left[\beta \sum_{j=1}^N S_j(i) S_j(i+1) + \ln A[S(i)] \right] \\
&\quad + \frac{nN}{2} \ln [(1-p)p] \tag{2.13}
\end{aligned}$$

ya que

$$H = \sum_{i=0}^{n-1} h[S(i), S(i+1)]$$

La estructura de la población después de n generaciones corresponde a la estructura de la superficie en el bloque de longitud n . El límite cuando $n \rightarrow \infty$ proporciona la distribución estacionaria del sistema que corresponde a la superficie del bloque infinito. En este punto, el comportamiento es independiente de las condiciones iniciales del sistema como la temperatura.

Tarazona [5] estudió la distribución estacionaria de la población en diferentes paisajes de aptitud. Obtuvo soluciones numéricas por medio de métodos usados en la física que no se discutirán en este trabajo. Sin embargo, se analizarán los resultados que obtuvo. El paisaje de aptitud más simple que estudió corresponde al de la secuencia maestra $S_0 = (\xi_1, \xi_2, \dots, \xi_N)$ donde $A[S_0] = A_0$ y $A[S] = A_1 < A_0 \quad \forall S \neq S_0$. La figura 2.14 muestra la distribución estacionaria de la población para este paisaje de aptitud.

Los datos corresponden a un sistema donde $S_0 = (1, 1, \dots, 1)$, $N = 20$ y $A_0/A_1 = 10$. La gráfica 2.14 muestra la concentración relativa de las cuerdas en el punto estable como función de la tasa de mutación p . Los números asociados con cada curva indican la distancia de Hamming respecto a la secuencia maestra. Obsérvese que la distribución cuasi-especie cambia suavemente cuando se aproxima al umbral de error. Cuando se da este comportamiento se dice que la transición de fase es de **segundo orden**. El umbral de error se da en $p^* = 0.11$. Después del umbral de error se tiene una distribución no localizada y la secuencia maestra tiene una concentración muy baja en la población. Por otro lado, obsérvese que hay más cuerdas con distancia de Hamming 9 que con distancia de Hamming 6 cuando $p > 0.11$. Esto se debe a que en una población aleatoria hay ${}^{20}C_9$ y ${}^{20}C_6$ cuerdas con distancia de Hamming 9 y 6 respectivamente donde ${}^N C_j = \frac{N!}{j!(N-j)!}$. Además, ${}^{20}C_9 = {}^{20}C_{11}$ por lo que hay el mismo número

de cuerdas con distancia de Hamming 9 y 11 en una población aleatoria para cuerdas de tamaño 20.

Al principio de este capítulo se mostró que otra manera de estudiar las transiciones de fase es mediante el parámetro de orden que en este caso se define como:

$$m = \frac{1}{N} \sum_{j=1}^N \xi_j \langle S_j \rangle$$

donde $\langle S_j \rangle$ es el promedio del valor del j -ésimo *bit* en toda la población y ξ es la secuencia maestra. El parámetro de orden se calcula con la población correspondiente a la distribución estacionaria, es decir, con la estructura de la superficie del bloque de longitud n . Supóngase que se tiene la siguiente población en la distribución estacionaria $(1, 1, 1)$, $(1, -1, 1)$, $(-1, -1, 1)$ y $(-1, -1, -1)$. En este caso, $m = \frac{1}{3}[0 - \frac{2}{4} + \frac{2}{4}] = 0$. Obsérvese que $-1 \leq m \leq 1$. Si $m = 1$, la población está formada únicamente por la secuencia maestra y por lo tanto está totalmente ordenada. Si $m = 0$, se tiene una población aleatoria, es decir, una población totalmente desordenada. Por último, si $m = -1$, la población está formada por el inverso de la secuencia maestra, es decir, por la cuerda $(-1, -1, \dots, -1)$.

La gráfica 2.15 muestra el parámetro de orden en la distribución estacionaria como función de la tasa de mutación p para el modelo anterior. El parámetro de orden se aproxima a cero suavemente. Al igual que la gráfica 2.14, se tiene una población aleatoria a partir de $p = 0.11$.

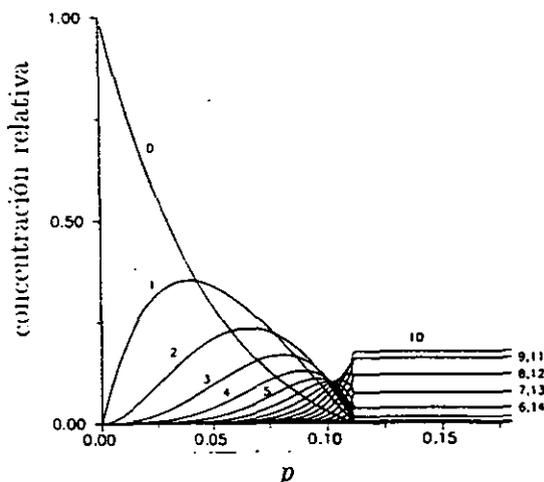


Figura 2.14: Distribución estacionaria de la población para el paisaje de aptitud de la secuencia maestra.

Tarazona considera un segundo paisaje de aptitud con 2 máximos, uno en $S_0 = (\xi_1, \xi_2, \dots, \xi_N)$ y otro en $S_N = (-\xi_1, -\xi_2, \dots, -\xi_N)$. S_0 corresponde al pico más alto del paisaje de aptitud (máximo global) y S_N es menos alto pero más ancho. Es decir, $A[S_0] = A_0$, $A[S_N] = A_N$, $A[S_{N-1}] = A_{N-1}$ y para cualquier otra cuerda S , $A[S] = A_1 < A_{N-1} < A_N < A_0$. El parámetro de orden se define de la misma manera en este caso. Si $m = 1$ la población está formada por el máximo S_0 y si $m = -1$, se tiene el máximo local S_N . La gráfica 2.16 muestra el parámetro de orden del modelo anterior donde $N = 20$, $\frac{A_0}{A_1} = 10$, $\frac{A_N}{A_1} = 9.9$ y $\frac{A_{N-1}}{A_1} = 2$. Para tasas de mutación pequeñas, la distribución cuasi-especie está centrada en S_0 . A medida que aumenta la tasa de mutación, la distribución cuasi-especie se centra en S_N . Obsérvese que la transición de una distribución cuasi-especie a otra se da de manera repentina, por lo que la transición de fase es de primer orden. El umbral de error se da en $p^* = 0.113$.

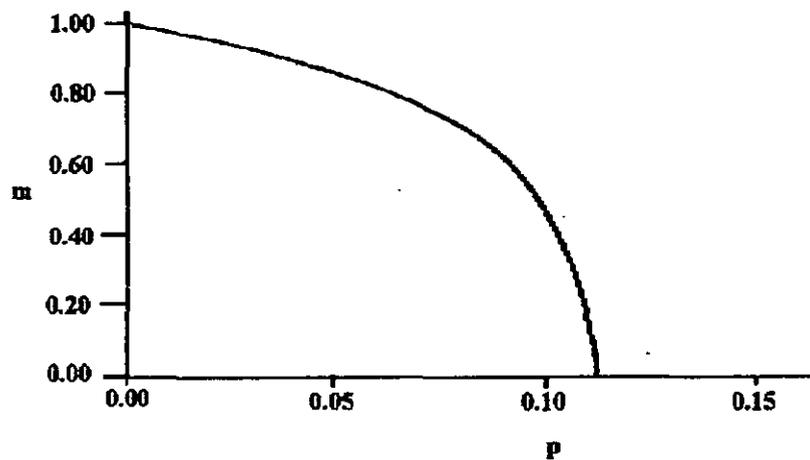


Figura 2.15: Parámetro de orden para el paisaje de aptitud de la secuencia maestra.

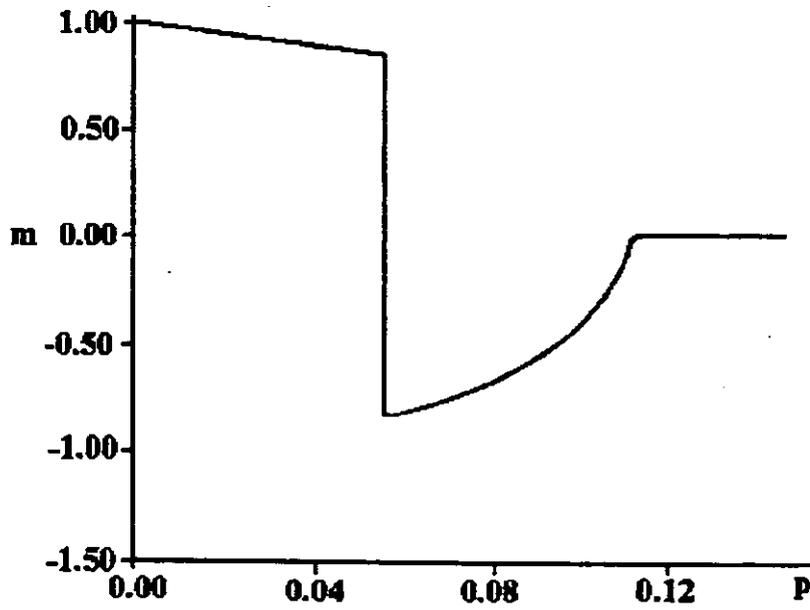


Figura 2.16: Parámetro de orden para el paisaje de aptitud con 2 picos.

CAPÍTULO 3

Resultados Experimentales de Transiciones de Fase.

En este capítulo, se estudian las transiciones de fase de primer orden y segundo orden en *AGs*. Los paisajes de aptitud elegidos son el de la secuencia maestra y el paisaje de aptitud con 2 picos que fueron estudiados por Tarazona. Los resultados experimentales, donde se tienen poblaciones finitas, se comparan con los obtenidos por Tarazona analíticamente. Por otro lado, se determina cómo varía el umbral de error como función de los parámetros del *AG*. Para ello, se eligió el paisaje de aptitud de la secuencia maestra ya que es muy simple y por lo tanto se pueden observar claramente los efectos que tienen sobre el umbral de error parámetros como el tamaño de la población, la presión selectiva del óptimo y el tamaño de la cuerda.

Para implementar el *AG* se deben fijar ciertos parámetros. Para el paisaje de aptitud de la secuencia maestra, se eligió el óptimo en $S_0 = (1, 1, \dots, 1)$ con aptitud de $A[S_0] = 10$. Las demás cuerdas tienen una aptitud de una unidad. En el caso del paisaje de aptitud con 2 picos, los óptimos locales se eligieron en $S_0 = (1, 1, \dots, 1)$, $S_N = (-1, -1, \dots, -1)$ y $S_{N-1} = (-1, -1, \dots, -1, 1)$ con $A[S_0] = 10$, $A[S_N] = 9.9$, $A[S_{N-1}] = 2$ y $A[S_i] = 1 \forall i \neq N, N-1, 0$. Dado que se desea estudiar el efecto de selección y mutación, no se consideró el operador de entrecruzamiento. El *AG* se corrió una vez para cada experimento. Los experimentos se realizaron con diferentes tamaños de población n . En la población inicial se incluyeron 10 secuencias maestras y las demás cuerdas se generaron aleatoriamente. Es importante tener la secuencia maestra en la población inicial porque de lo contrario el *AG* no encontraría el óptimo si la probabilidad de mutación es cero. La distribución estacionaria de la población se determinó empíricamente en 50 generaciones. La selección se implementó con ruleta.

La gráfica 3.1 muestra la concentración relativa de las cuerdas (respecto a la distancia de Hamming) como función de la tasa de mutación p . El tamaño de la cuerda se fijó en $N = 20$ para poder comparar estos resultados con los de

Tarazona. La población es de $n = 1724$ individuos. Para representar claramente los resultados se omitieron las curvas correspondientes a las distancias de Hamming 10, \dots , 20. Los resultados que se obtienen en esta gráfica concuerdan con los obtenidos con Tarazona ya que el umbral de error se alcanza en $p^* = 0.11$. La gráfica 3.2 muestra el parámetro de orden m correspondiente en el cual se puede observar que el umbral de error se da aproximadamente en $p^* = 0.11$.

A continuación, se estudia el efecto que tiene el tamaño finito de la población sobre el umbral de error. Para ello se consideraron cuerdas de tamaño $N = 10$. El número total de estados es $2^{10} = 1024$ que es 2^{10} veces menor al caso anterior donde $N = 20$. Esto permite variar el tamaño de la población (con respecto al número total de estados) con mayor facilidad en las simulaciones. Las gráficas 3.4, 3.5 y 3.8 muestran la concentración relativa de la población como función de la tasa de mutación para $n = 100$, $n = 1024$ y $n = 3000$ respectivamente. Las gráficas 3.3, 3.6 y 3.7 representan el parámetro de orden correspondiente a las gráficas 3.4, 3.5 y 3.8 respectivamente. Cuando el tamaño de la población es mucho menor al número de estados posibles (gráfica 3.4), las curvas presentan muchas fluctuaciones debido al efecto de tamaño finito. A medida que se incrementa el tamaño de la población, las fluctuaciones disminuyen.

Con el objeto de estudiar cómo varía el umbral de error como función de la presión selectiva del óptimo, se implementó la selección con torneo de tamaño 2, 10 y 20. Los resultados se muestran en las gráficas 3.9, 3.10 y 3.11 respectivamente. Obsérvese que entre mayor es el tamaño del torneo, el umbral de error aumenta. Además con torneo de tamaño 10 y con ruleta, el umbral de error coincide. Esto se debe a lo siguiente. Considérese una población inicial constituida por cada uno de los 2^{20} estados posibles. Es decir, $n(S_i, 0) = 1 \forall i$. Si se usa ruleta, el número esperado de secuencias maestras en la primera generación es:

$$\begin{aligned} n(S_0, 1) &= \frac{A[S_0]n(S_0, 0)}{\sum_i A[S_i]n(S_i, 0)} * n \\ &= \frac{10(1)}{\sum_i A[S_i]n(S_i, 0)} * n \end{aligned}$$

Por otro lado, el número esperado de cuerdas diferentes a la secuencia maestra en la primera generación es:

$$\begin{aligned} n(S_j, 1) &= \frac{A[S_j]n(S_j, 0)}{\sum_i A[S_i]n(S_i, 0)} * n \quad \text{con } j \neq 0 \\ &= \frac{1(1)}{\sum_i A[S_i]n(S_i, 0)} * n \quad \text{con } j \neq 0 \end{aligned}$$

Es decir, se espera que en la primera generación el número de secuencias maestras sea 10 veces más que el número esperado de cualquier otra cuerda en la población.

Si se considera torneo de tamaño 2 como operador de selección y la misma población inicial entonces en la primera generación habrá 2 secuencias maestras. Esto se debe a que en cada etapa del torneo se eligen 2 individuos aleatoriamente sin reemplazo. Las demás cuerdas pueden desaparecer o ser seleccionadas 1 ó 2 veces. Por lo tanto, en la primera generación, el número de secuencias maestras aparece en el mejor de los casos 2 veces más que cualquier otra cuerda. Dado que el tamaño de la población es fijo, esto significa que el número esperado de secuencias maestras en la primera generación (y por lo tanto en las subsecuentes generaciones) es menor si se usa torneo de tamaño 2. Como consecuencia, es más fácil perder la secuencia maestra por mutaciones y por lo tanto, el umbral de error se alcanza para una tasa de mutación menor. Cuando el tamaño del torneo es de 10, la secuencia maestra tiene la misma ventaja selectiva que con ruleta y por lo tanto el umbral de error coincide. Si el tamaño del torneo es mayor que 10, la secuencia maestra tiene mayor presión selectiva que en el caso de ruleta y por lo tanto el umbral de error se alcanza para una tasa de mutación mayor.

Otro parámetro que se consideró para estudiar cómo se modifica el umbral de error es el tamaño de la cuerda N . Dado que la tasa de mutación se aplica a cada *bit* de la cuerda, las cuerdas de mayor tamaño tienen más probabilidades de mutar para un p fijo. Por lo tanto, si N es grande es más probable que la secuencia maestra mute a otra cuerda. Entonces, se espera que entre mayor sea el tamaño de la cuerda N , menor sea el umbral de error. A continuación se obtiene una aproximación del umbral de error como función del tamaño de la cuerda para el paisaje de aptitud de la secuencia maestra. El número esperado de secuencias maestras en la generación $t + 1$ si se usa ruleta es:

$$n(S_0, t + 1) = \frac{1}{\bar{A}(t)} \sum_i A[S_i] n(S_i, t) (1 - p)^{N - d^H(S_i, S_0)} p^{d^H(S_i, S_0)}$$

donde $\bar{A}(t)$ es la aptitud promedio de la población en la generación t . Si el número de estados posibles del sistema es muy grande, es razonable suponer que una cuerda mutada no puede volver a formar la secuencia maestra. Es decir, se desprecian las mutaciones "hacia atrás". Entonces,

$$n(S_0, t + 1) = \frac{1}{\bar{A}(t)} A[S_0] n(S_0, t) (1 - p)^N$$

Por otro lado,

$$\bar{A}(t) = \sum_i A[S_i] P(S_i, t) = A[S_0] P(S_0, t) + A[S] (1 - P(S_0, t))$$

donde $A[S]$ es la aptitud de las cuerdas diferentes a la secuencia maestra.

$$\Rightarrow n(S_0, t + 1) = \frac{A[S_0]n(S_0, t)(1 - p)^N}{A[S_0]P(S_0, t) + A[S](1 - P(S_0, t))}$$

En la distribución estacionaria, $n(S_0, t + 1) = n(S_0, t)$

$$\Rightarrow 1 = \frac{A[S_0](1 - p)^N}{A[S_0]P(S_0, t) + A[S](1 - P(S_0, t))}$$

En el umbral de error p^* , el número de secuencias maestras es muy pequeño por lo que $P(S_0, t)$ se puede despreciar. De esta manera se obtiene:

$$\begin{aligned} 1 &= \frac{A[S_0]}{A[S]}(1 - p^*)^N \\ \Rightarrow \frac{A[S]}{A[S_0]} &= (1 - p^*)^N \\ \Rightarrow p^* &= 1 - \left(\frac{A[S]}{A[S_0]}\right)^{\frac{1}{N}} \end{aligned} \quad (3.1)$$

En el problema bajo consideración, $A[S] = 1$ y $A[S_0] = 10$ entonces

$$p^* = 1 - \left(\frac{1}{10}\right)^{\frac{1}{N}}$$

La figura 3.12 muestra la gráfica de esta función. En esta gráfica también se presenta el umbral de error como función de N obtenido experimentalmente. Obsérvese que esta curva es muy similar a la obtenida teóricamente.

Por último, la gráfica 3.13 muestra el parámetro de orden para el paisaje de aptitud con 2 picos. En este caso, el tamaño de la cuerda se fijó en $N = 10$. Para tener una cuerda de cada estado en la población inicial, el tamaño de la población es entonces de $2^{10} = 1024$ individuos. La selección se implementó con torneo de tamaño 2 para evitar que se perdiera la secuencia maestra. Cualitativamente, se obtienen los mismos resultados que Tarazona. Se observa una transición de fase de primer orden. La transición de la distribución cuasi-especie alrededor de la secuencia maestra S_0 a la distribución cuasi-especie centrada en el óptimo local S_N se da de manera repentina aproximadamente en $p = 0.065$. El umbral de error se alcanza aproximadamente en $p^* = 0.085$.

Los paisajes de aptitud considerados en este capítulo son muy simples pero es por esto que los efectos de los operadores genéticos se pueden determinar claramente. De acuerdo con los resultados obtenidos se tienen las siguientes conclusiones. Se observó que el umbral de error es menor cuando la presión selectiva del óptimo disminuye. Entonces, entre menor sea la presión selectiva del óptimo, menor debe ser la tasa de mutación que se fija en el *AG*. Por otro lado, dado que el umbral de error es aproximadamente inversamente proporcional al tamaño de la cuerda; la tasa de mutación debe ser menor entre más grande sea el tamaño de la cuerda. También se observó que las fluctuaciones disminuyen a medida que aumenta el tamaño de la población. Por lo tanto, es conveniente trabajar en las simulaciones con poblaciones grandes con respecto al número total de estados del sistema para observar el comportamiento de éste de manera más clara.

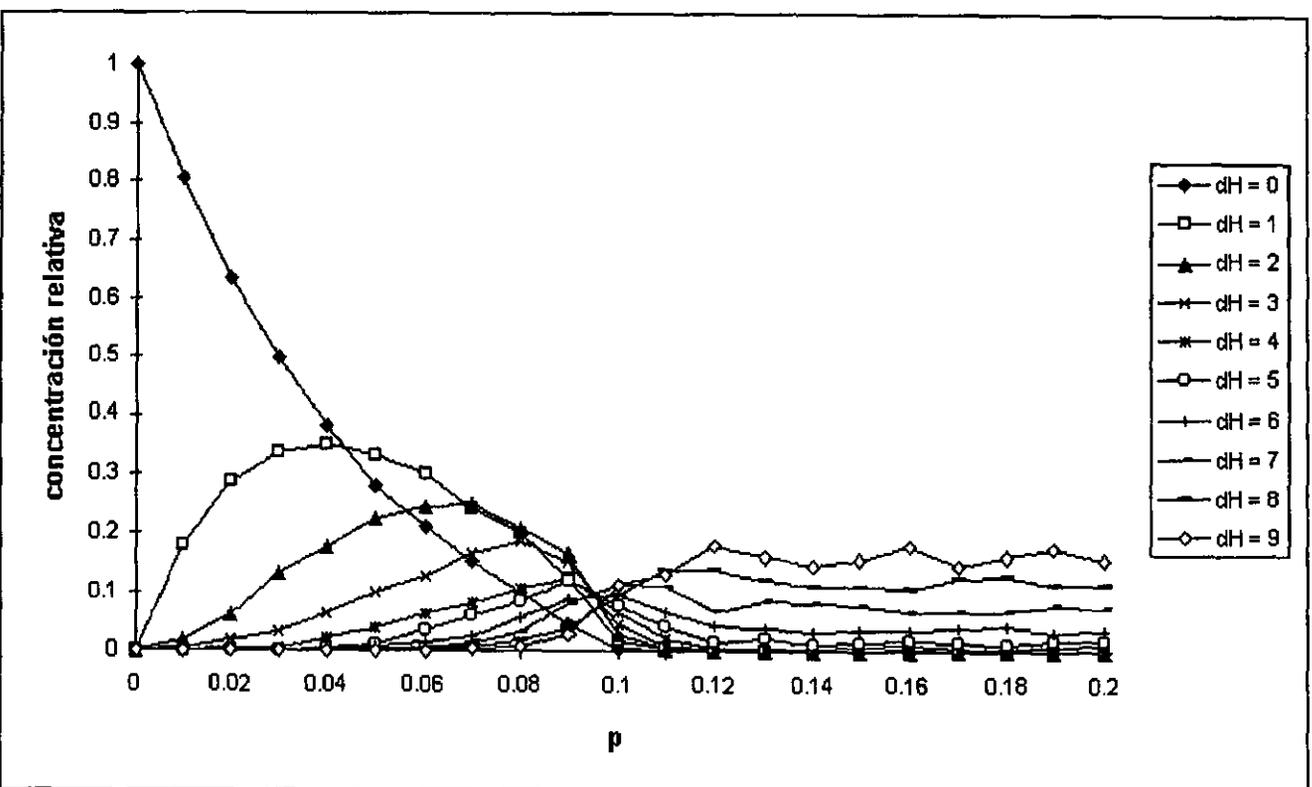


Figura 3.1: Distribución estacionaria de la población para el paisaje de aptitud de la secuencia maestra donde $n = 1724$.

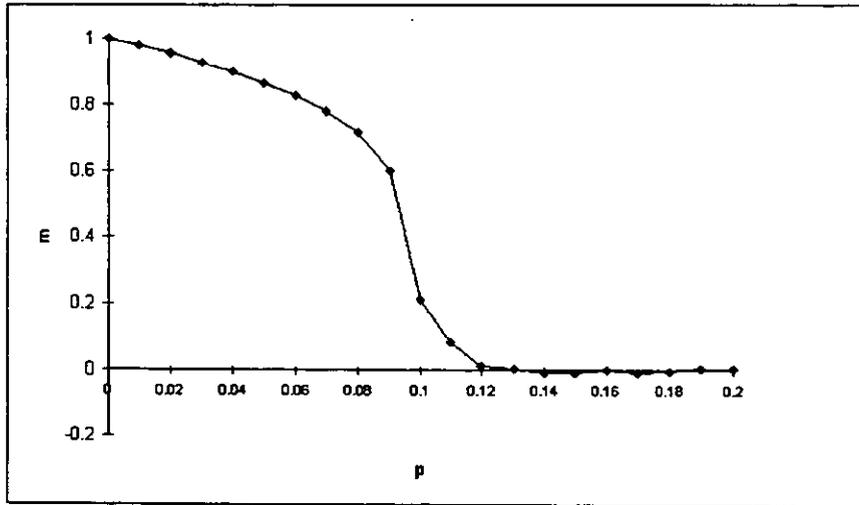


Figura 3.2: Parámetro de orden en el paisaje de aptitud de la secuencia maestra con $n = 1724$.

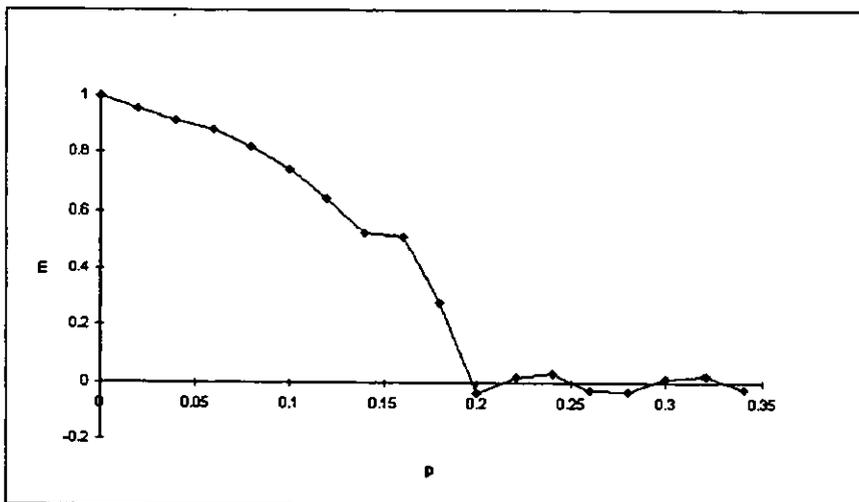


Figura 3.3: Parámetro de orden en el paisaje de aptitud de la secuencia maestra con $n = 100$ y $N = 10$.

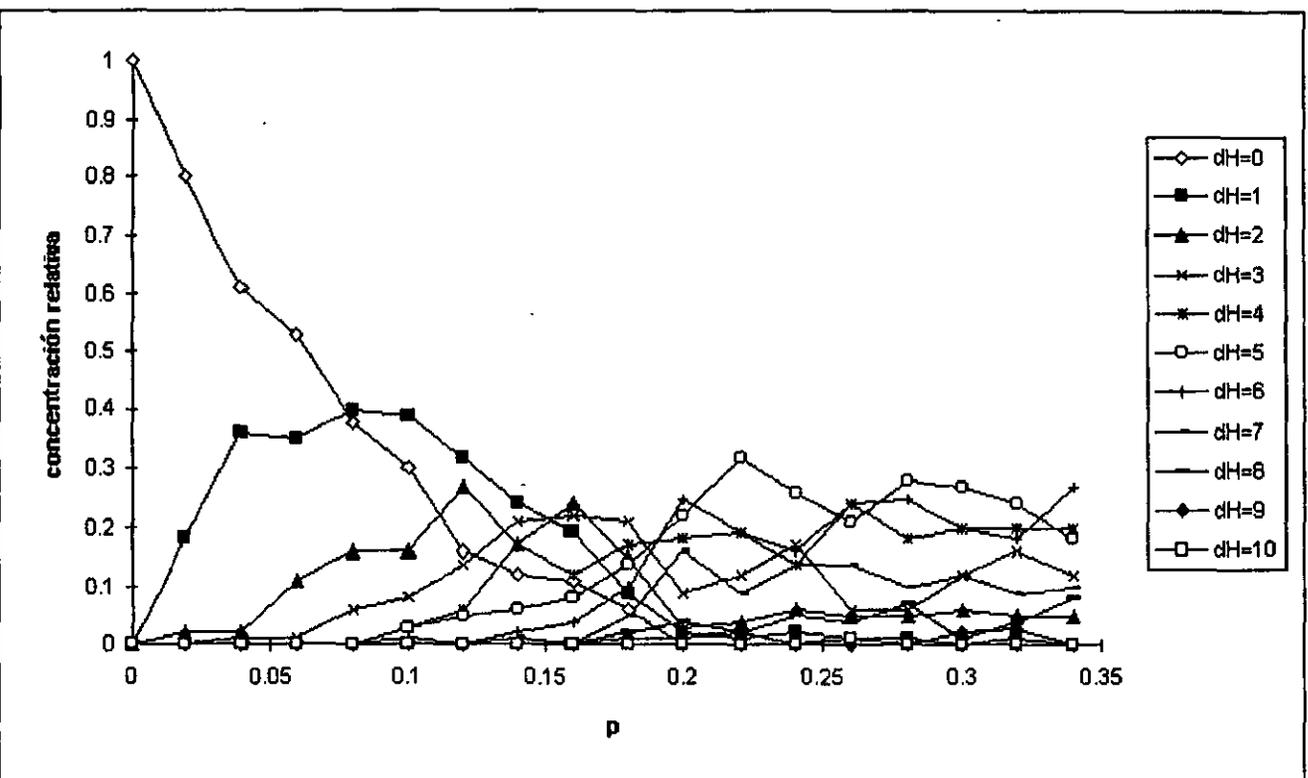


Figura 3.4: Distribución estacionaria de la población para el paisaje de aptitud de la secuencia maestra donde $n = 100$ y $N = 10$.

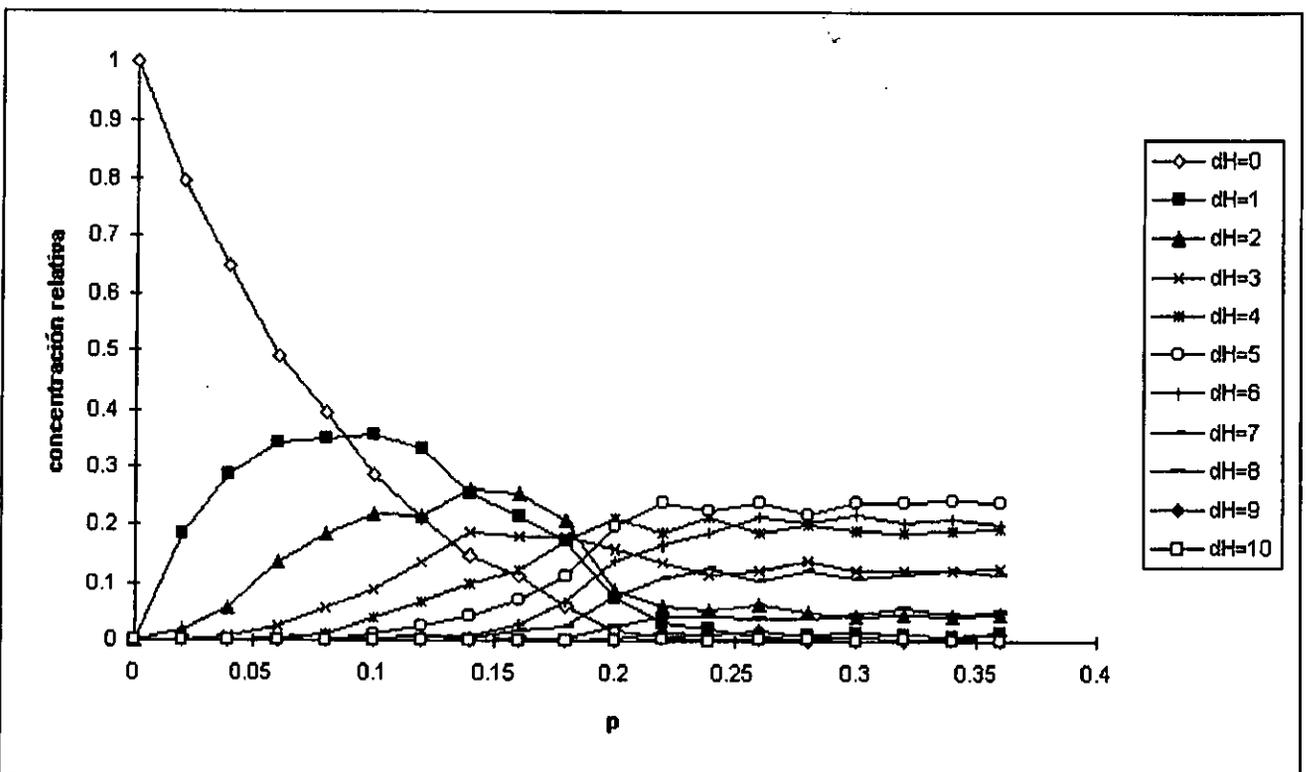


Figura 3.5: Distribución estacionaria de la población para el paisaje de aptitud de la secuencia maestra donde $n = 1024$ y $N = 10$.

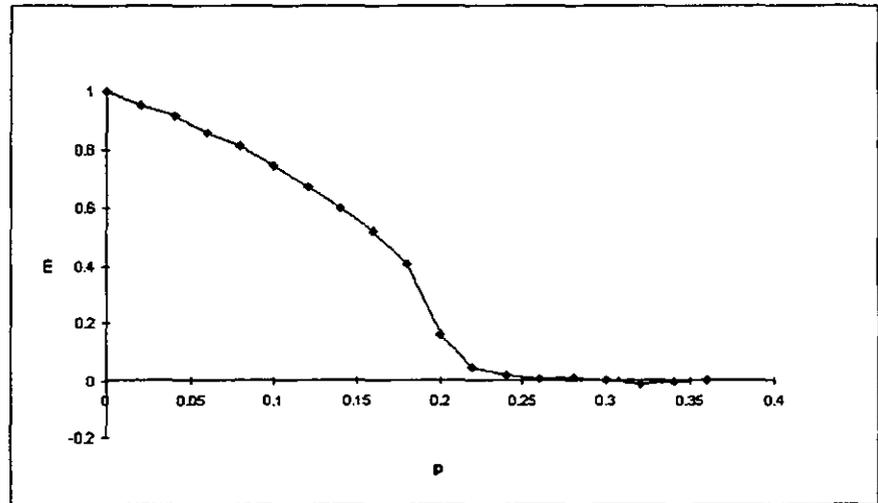


Figura 3.6: Parámetro de orden en el paisaje de aptitud de la secuencia maestra con $n = 1024$ y $N = 10$.

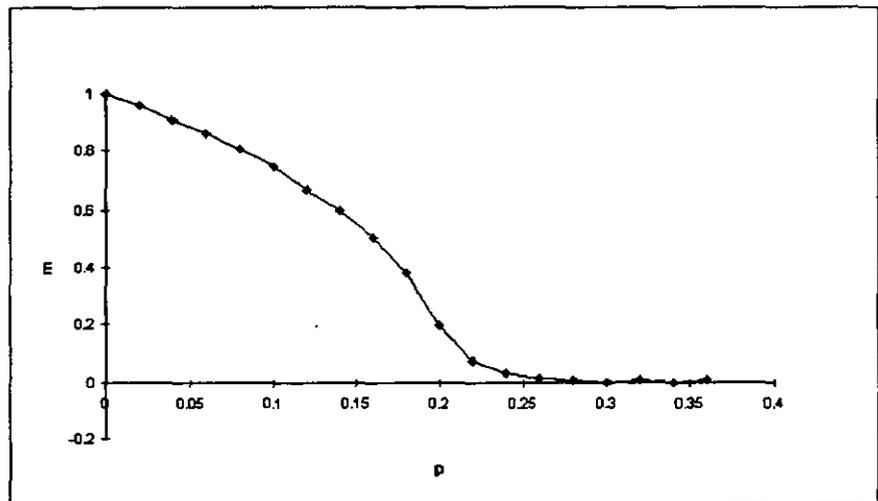


Figura 3.7: Parámetro de orden en el paisaje de aptitud de la secuencia maestra con $n = 3000$ y $N = 10$.

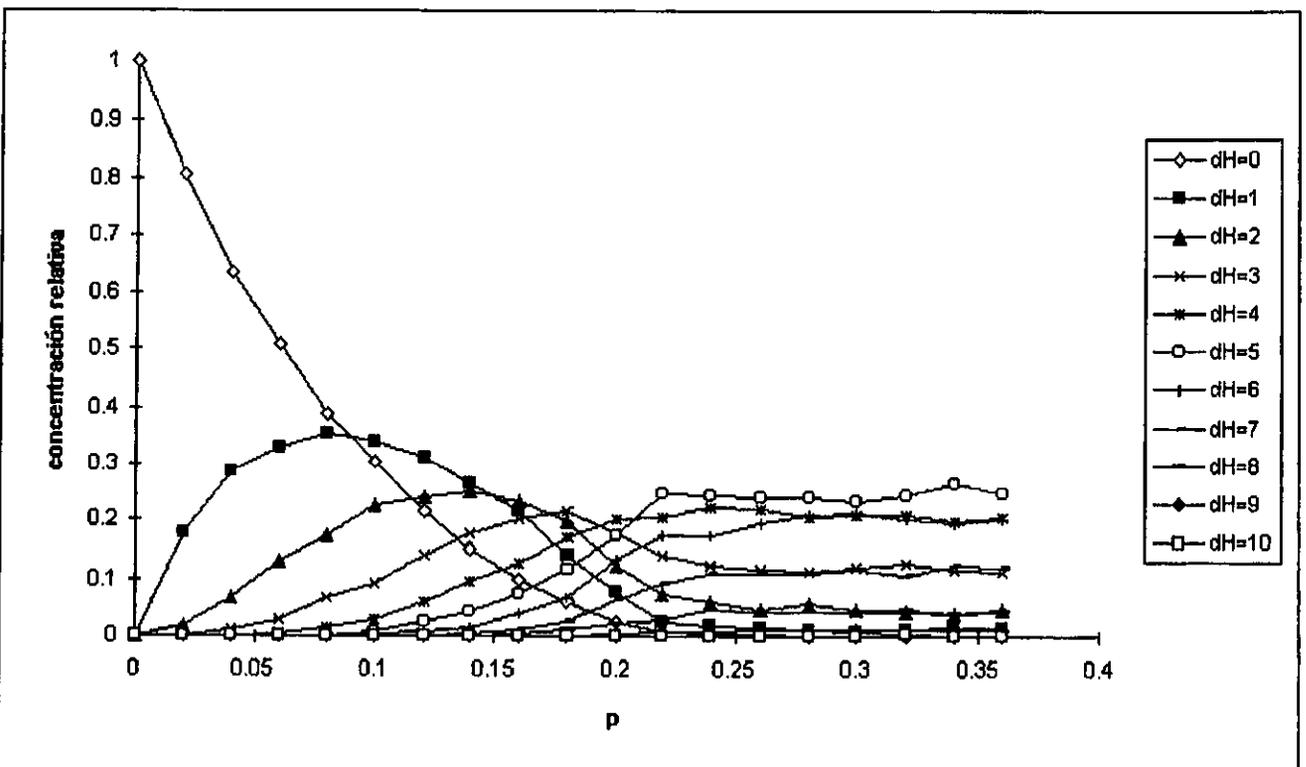


Figura 3.8: Distribución estacionaria de la población para el paisaje de aptitud de la secuencia maestra donde $n = 3000$ y $N = 10$.

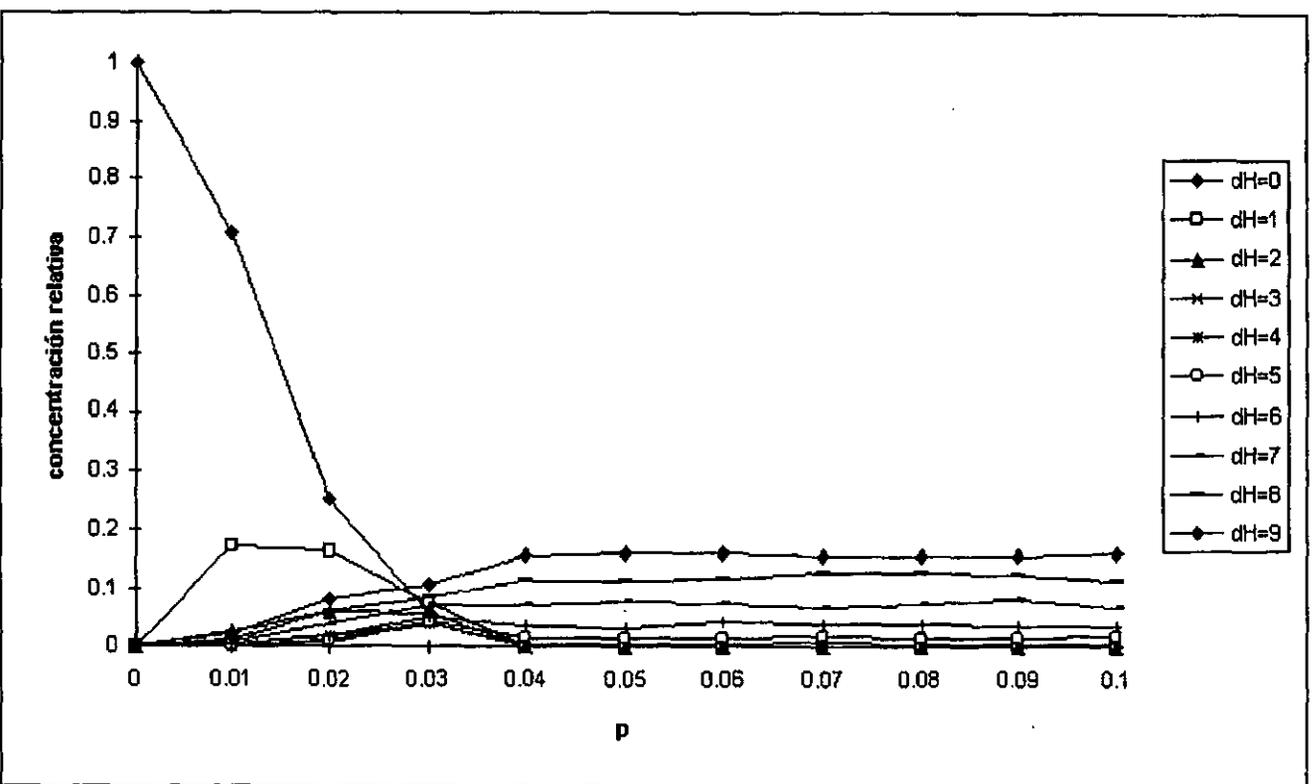


Figura 3.9: Distribución estacionaria de la población para el paisaje de aptitud de la secuencia maestra con torneo de tamaño 2.

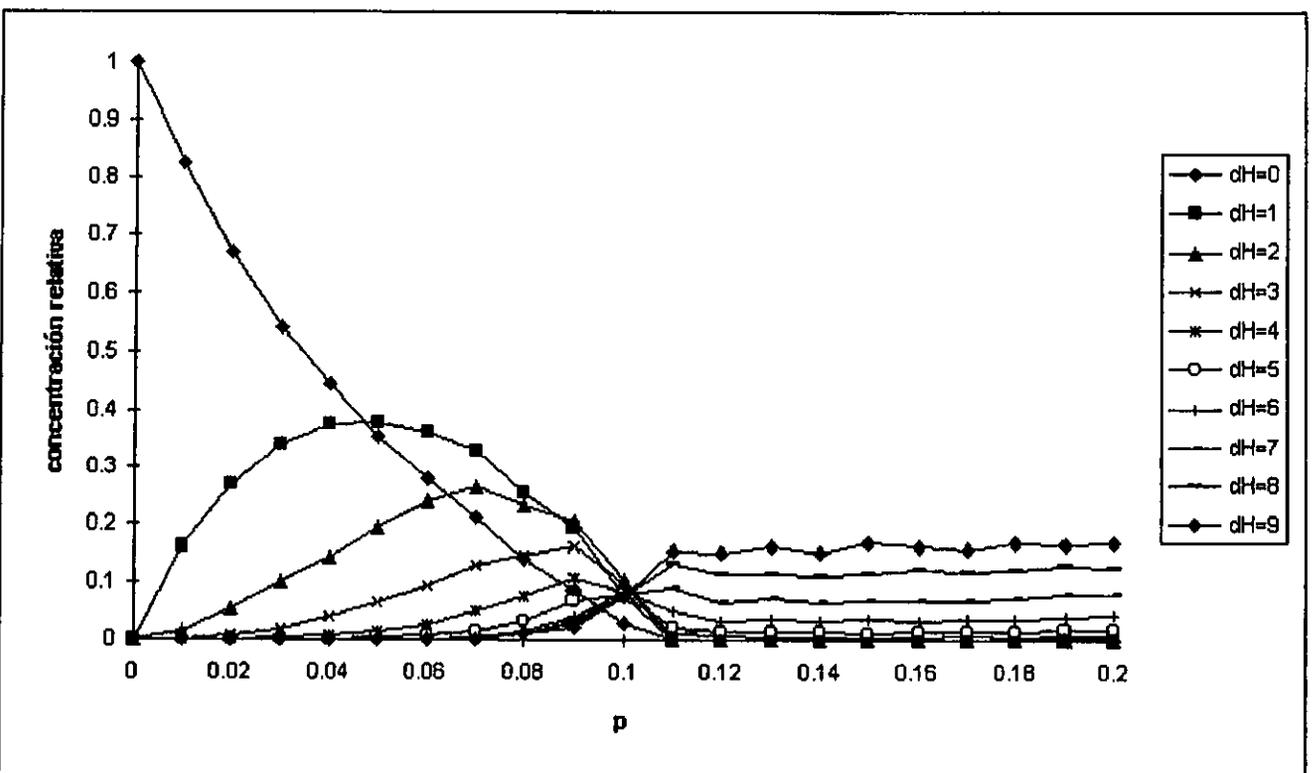


Figura 3.10: Distribución estacionaria de la población para el paisaje de aptitud de la secuencia maestra con torneo de tamaño 10.

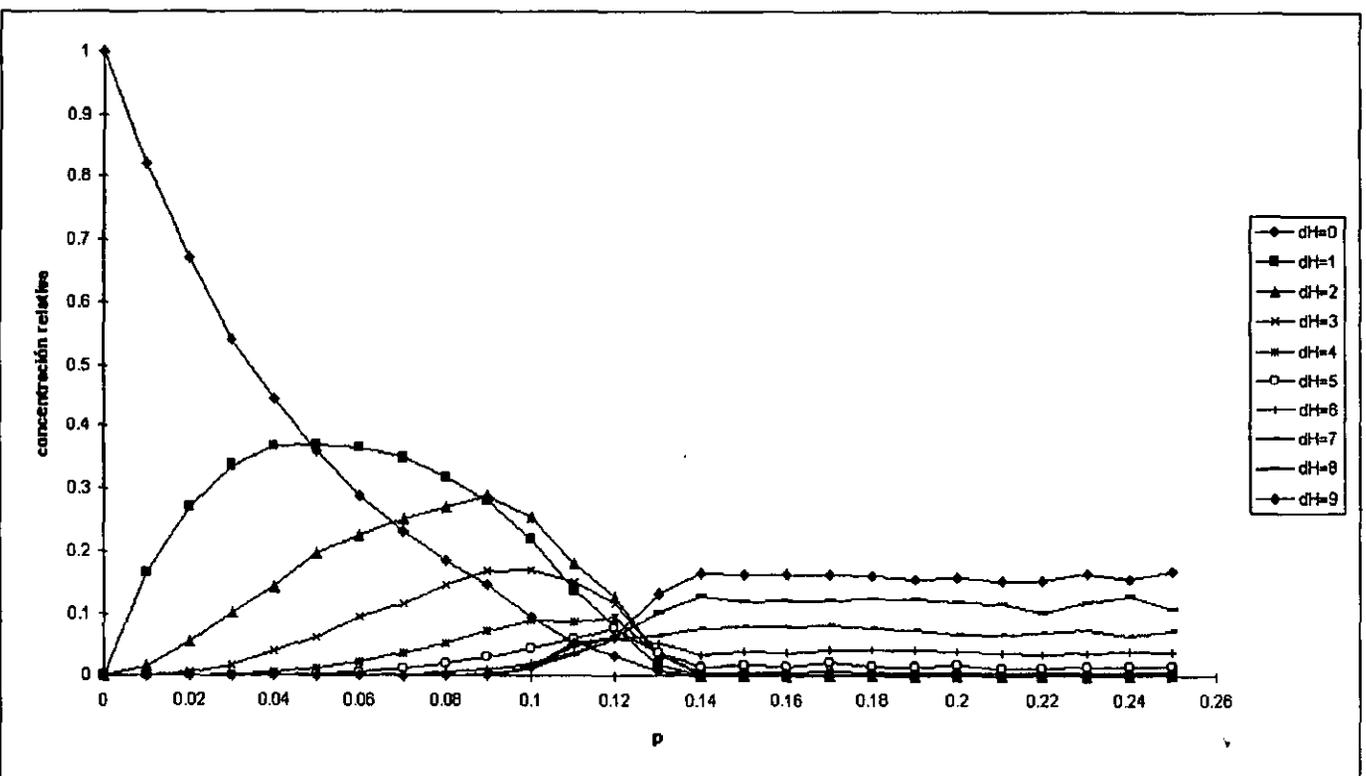


Figura 3.11: Distribución estacionaria de la población para el paisaje de aptitud de la secuencia maestra con torneo de tamaño 20.

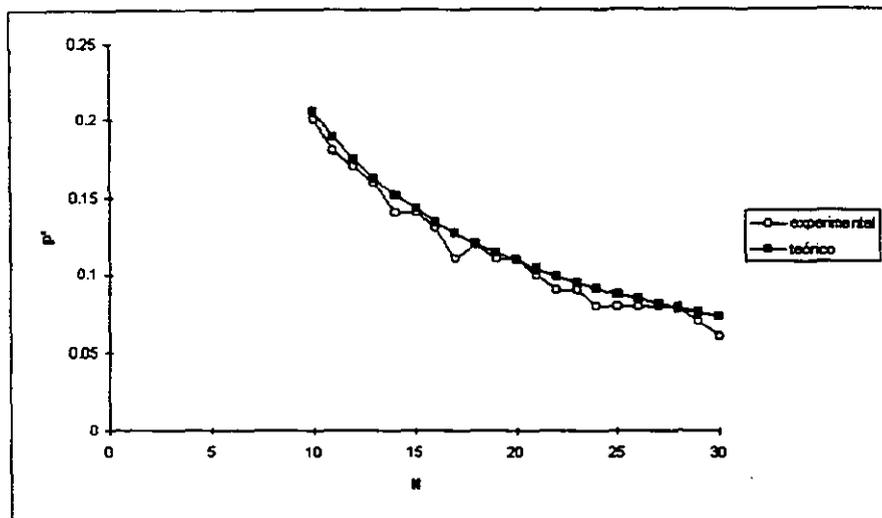


Figura 3.12: Umbral de error como función del tamaño de la cuerda N .

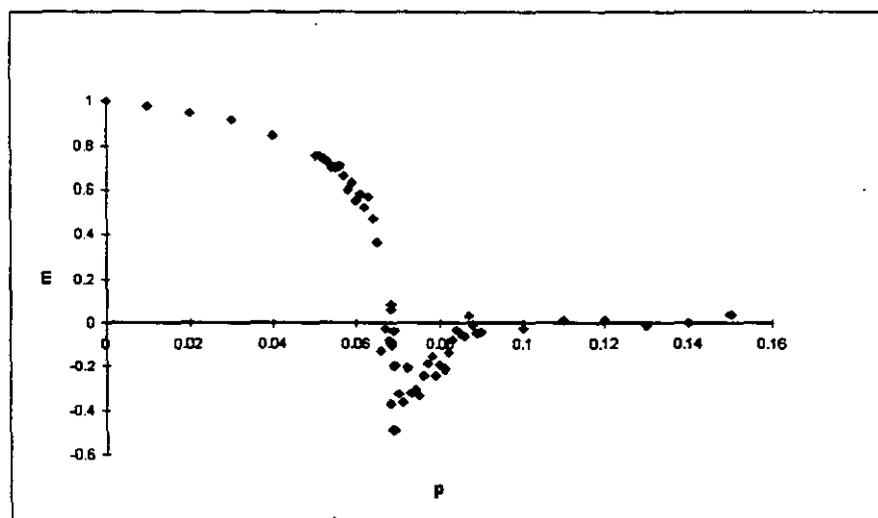


Figura 3.13: Parámetro de orden para el paisaje de aptitud plano con 2 picos.

CAPÍTULO 4

Grados de Libertad Efectivos.

El objetivo de este capítulo es caracterizar los grados de libertad efectivos de un AG , es decir, se determinan las características de los esquemas que construyen soluciones aptas. Para ello, se simplifica la ecuación de evolución para cuerdas con el objeto de mostrar que entrecruzamiento introduce la noción de esquemas. Además, se analiza la forma en cómo entrecruzamiento combina pedazos de soluciones existentes (esquemas) para encontrar soluciones aptas, es decir, se discute la idea de bloques constructores. Esto permitirá caracterizar los grados de libertad efectivos en un AG . Por último, se introduce el concepto de aptitud efectiva el cual da una mejor idea de la evolución de esquemas en el tiempo.

4.1 Ejemplo de un AG con Selección y Cruza.

Con el objeto de ilustrar el papel que tiene el operador de entrecruzamiento en un AG , se resolverá un problema sencillo. Sea $f(x) = x^2$ con $x \in [0, 31]$. Para codificar las variables de este problema, se usa una codificación binaria. A continuación, se muestra una iteración del AG que utiliza los operadores de selección y entrecruzamiento.

cuerda	pob. inicial	x	$f(x)$	$P'(c_i, 1)$	$n(c_i, 1)$
c_1	01101	13	169	0.144	0.58
c_2	11000	24	576	0.4923	1.97
c_3	01000	8	64	0.0547	0.22
c_4	10011	9	361	0.3085	1.23
		suma	1170	1	4
		promedio	292.5	0.25	1
		máximo	576	0.423	1.97

Esta tabla muestra el número esperado de cuerdas c_i que se obtienen con se-

lección. La siguiente tabla proporciona la nueva población después de aplicar selección y entrecruzamiento.

cuerda	# de veces seleccionada	pob. después de selección	parejas seleccionadas al azar
C_1	1	01101	2
C_2	2	11000	1
C_3	0	11000	4
C_4	1	10011	3

punto de cruza	pob. después de cruza	x	$f(x)$
4	11001	25	625
4	01100	12	144
2	11011	27	729
2	10000	16	256
		suma	1754
		promedio	438.5 (> 292.5)
		máximo	729

De estas tablas se pueden hacer algunas observaciones. Primero, las cuerdas 11000 y 01000 sólo difieren en el primer *bit* y sin embargo la aptitud de cada una de ellas es muy diferente. Además, en la segunda generación, aumenta el número de cuerdas cuyo primer *bit* vale uno. Esto hace suponer que es importante que el primer *bit* valga uno para que la cuerda tenga una aptitud alta. Es decir, las cuerdas asociadas con el esquema 1**** tienen una aptitud alta. Nótese que una cuerda está asociada con 2^5 esquemas por lo que cada vez que se considera una cuerda se analizan 2^5 esquemas. Esto significa que un AG realiza una búsqueda en paralelo.

Por otro lado, la cuerda de mayor aptitud después de la primera generación es 11011, la cual se obtuvo al entrecruzar las 2 cuerdas más aptas de la primera generación. La cuerda resultante tiene una aptitud mayor que las originales. Es decir, entrecruzamiento construye cuerdas aptas al combinar pedazos de cuerdas existentes. En las siguientes secciones se formalizan estos conceptos.

4.2 Ecuación de Evolución para Esquemas.

En el capítulo 2 se desarrolló la siguiente ecuación de evolución para cuerdas en un AG con los operadores de selección y entrecruzamiento.

$$\begin{aligned}
P(c_i, t+1) &= P'(c_i, t) \\
&- \frac{p_c}{N-1} \sum_{C_j \neq C_i} \sum_{k=1}^{N-1} C_{C_i, C_j}^{(1)}(k) P'(c_i, t) P'(c_j, t) \\
&+ \frac{p_c}{N-1} \sum_{C_j \neq C_i} \sum_{C_l \neq C_i} \sum_{k=1}^{N-1} C_{C_j, C_l}^{(2)}(k) P'(c_j, t) P'(c_l, t) \quad (4.1)
\end{aligned}$$

Para simplificar esta ecuación se hace lo siguiente. Sea

$$\begin{aligned}
A &= \{c_j : d_L^H(i, j) > 0 \text{ y } d_R^H(i, j) > 0\} \quad \text{y} \\
B &= \{c_j : (d_L^H(i, j) > 0 \text{ y } d_R^H(i, j) = 0) \text{ ó } (d_L^H(i, j) = 0 \text{ y } d_R^H(i, j) > 0)\}
\end{aligned}$$

es decir, A es el conjunto de cuerdas c_j que difieren de c_i en ambos lados del punto de cruce k y B es el conjunto de cuerdas c_j que difieren de c_i en sólo un lado del punto de cruce k . Entonces, el término destructivo de esta ecuación se puede escribir como:

$$\begin{aligned}
\sum_{C_j \neq C_i} C_{C_i, C_j}^{(1)}(k) P'(c_i, t) P'(c_j, t) &= \sum_{C_j \in A} P'(c_i, t) P'(c_j, t) \\
&= \sum_{C_j} P'(c_i, t) P'(c_j, t) \\
&- \sum_{C_j \in B} P'(c_i, t) P'(c_j, t) \\
&- P'(c_i, t) P'(c_i, t)
\end{aligned}$$

$$\text{pero } \sum_{C_j} P'(c_i, t) P'(c_j, t) = P'(c_i, t) \sum_{C_j} P'(c_j, t) = P'(c_i, t)$$

$$\begin{aligned}
\Rightarrow \sum_{C_j \neq C_i} C_{C_i, C_j}^{(1)}(k) P'(c_i, t) P'(c_j, t) &= P'(c_i, t) \\
&- \sum_{C_j \in B} P'(c_i, t) P'(c_j, t) \\
&- P'(c_i, t) P'(c_i, t) \quad (4.2)
\end{aligned}$$

El término constructivo se puede reescribir de la siguiente manera:

$$\begin{aligned}
\sum_{C_j \neq C_i} \sum_{C_l \neq C_i} C_{C_j C_l}^{(2)}(k) P'(C_j, t) P'(C_l, t) &= \sum_{C_j} \sum_{C_l} C_{C_j C_l}^{(2)}(k) P'(C_j, t) P'(C_l, t) \\
&\quad - \sum_{C_j} C_{C_j C_i}^{(2)}(k) P'(C_j, t) P'(C_i, t) \\
&\quad - \sum_{C_l} C_{C_i C_l}^{(2)}(k) P'(C_i, t) P'(C_l, t) \\
&\quad - C_{C_i C_i}^{(2)}(k) P'(C_i, t) P'(C_i, t) \\
&= \underbrace{\sum_{C_j} \sum_{C_l} C_{C_j C_l}^{(2)}(k) P'(C_j, t) P'(C_l, t)}_{(I)} \\
&\quad - \underbrace{2 \sum_{C_j} C_{C_j C_i}^{(2)}(k) P'(C_j, t) P'(C_i, t)}_{(II)} \\
&\quad - \underbrace{C_{C_i C_i}^{(2)}(k) P'(C_i, t) P'(C_i, t)}_{(III)} \quad (4.3)
\end{aligned}$$

Se analizarán los términos (I), (II) y (III) por separado. Considérese el término (I). Dado que $C_j \neq C_i$ y $C_l \neq C_i$ se tiene que:

$$\begin{aligned}
C_{C_j C_l}^{(2)}(k) &= \frac{1}{2} \left[\delta(d_L^H(i, j)) \delta(d_R^H(i, l)) + \delta(d_R^H(i, j)) \delta(d_L^H(i, l)) \right] \\
&= \begin{cases} \frac{1}{2} & \text{si } (C_j \supset C_i^L \text{ y } C_l \supset C_i^R) \text{ ó } (C_j \supset C_i^R \text{ y } C_l \supset C_i^L) \\ 0 & \text{en otro caso} \end{cases}
\end{aligned}$$

Entonces

$$\begin{aligned}
\sum_{C_j} \sum_{C_i} C_{C_j C_i}^{(2)}(k) P'(C_j, t) P'(C_i, t) &= \frac{1}{2} \sum_{C_j \supset C_i^L} \sum_{C_i \supset C_j^R} P'(C_j, t) P'(C_i, t) \\
&+ \frac{1}{2} \sum_{C_j \supset C_i^R} \sum_{C_i \supset C_j^L} P'(C_j, t) P'(C_i, t) \\
&= \frac{1}{2} \sum_{C_j \supset C_i^L} \sum_{C_i \supset C_j^R} P'(C_j, t) P'(C_i, t) \\
&+ \frac{1}{2} \sum_{C_i \supset C_j^R} \sum_{C_j \supset C_i^L} P'(C_j, t) P'(C_i, t) \\
&= \sum_{C_j \supset C_i^L} \sum_{C_i \supset C_j^R} P'(C_j, t) P'(C_i, t)
\end{aligned}$$

En el término (II) se tiene que $C_j \neq C_i$ por lo tanto

$$\begin{aligned}
C_{C_j C_i}^{(2)}(k) &= \frac{1}{2} \left[\delta(d_L^H(i, j)) \delta(d_R^H(i, i)) + \delta(d_R^H(i, j)) \delta(d_L^H(i, i)) \right] \\
&= \frac{1}{2} \left[\delta(d_L^H(i, j)) + \delta(d_R^H(i, j)) \right] \\
&= \begin{cases} \frac{1}{2} & \text{si } C_j \in B \\ 0 & \text{si } C_j \in A \end{cases}
\end{aligned}$$

Es decir, $C_{C_j C_i}^{(2)}(k) = \frac{1}{2}$ si C_j es diferente de C_i en un sólo lado del punto de entrecruzamiento k . Entonces,

$$2 \sum_{C_j \in B} C_{C_j C_i}^{(2)}(k) P'(C_j, t) P'(C_i, t) = \sum_{C_j \in B} P'(C_j, t) P'(C_i, t)$$

Por último, considérese el término (III). Se tiene que

$$\begin{aligned}
C_{C_i C_i}^{(2)}(k) &= \frac{1}{2} \left[\delta(d_L^H(i, i)) \delta(d_R^H(i, i)) + \delta(d_R^H(i, i)) \delta(d_L^H(i, i)) \right] \\
&= 1
\end{aligned}$$

Por lo tanto,

$$C_{C_i C_i}^{(2)}(k) P'(C_i, t) P'(C_i, t) = P'(C_i, t) P'(C_i, t)$$

Entonces, el término "constructivo" (ec. 4.3) se simplifica de la siguiente manera:

$$\begin{aligned} \sum_{C_j \neq C_i} \sum_{C_l \neq C_i} C_{C_j C_l}^{(2)}(k) P'(C_j, t) P'(C_l, t) &= \sum_{C_j \supset C_i^L} \sum_{C_l \supset C_i^R} P'(C_j, t) P'(C_l, t) \\ &\quad - \sum_{C_j \in B} P'(C_j, t) P'(C_i, t) \\ &\quad - P'(C_i, t) P'(C_i, t) \end{aligned} \quad (4.4)$$

Por lo tanto, el término de entrecruzamiento que se obtiene a partir de las ecuaciones 4.2 y 4.4 es:

$$\begin{aligned} &-\frac{p_c}{N-1} \sum_{k=1}^{N-1} \left[P'(C_i, t) - \sum_{C_j \in B} P'(C_i, t) P'(C_j, t) - P'(C_i, t) P'(C_i, t) \right] \\ &+ \frac{p_c}{N-1} \sum_{k=1}^{N-1} \left[\sum_{C_j \supset C_i^L} \sum_{C_l \supset C_i^R} P'(C_j, t) P'(C_l, t) - \sum_{C_j \in B} P'(C_j, t) P'(C_i, t) \right. \\ &\quad \left. - P'(C_i, t) P'(C_i, t) \right] \\ &= \frac{-p_c}{N-1} \sum_{k=1}^{N-1} \left[P'(C_i, t) - \sum_{C_j \supset C_i^L} \sum_{C_l \supset C_i^R} P'(C_j, t) P'(C_l, t) \right] \end{aligned}$$

Entonces, la ecuación 4.1 se simplifica como:

$$\begin{aligned} P(C_i, t+1) &= P'(C_i, t) - \frac{p_c}{N-1} \sum_{k=1}^{N-1} \left[P'(C_i, t) \right. \\ &\quad \left. - \sum_{C_j \supset C_i^L} \sum_{C_l \supset C_i^R} P'(C_j, t) P'(C_l, t) \right] \\ \Rightarrow P(C_i, t+1) &= P'(C_i, t) - \frac{p_c}{N-1} \sum_{k=1}^{N-1} \left[P'(C_i, t) \right. \\ &\quad \left. - P'(C_i^L, t) P'(C_i^R, t) \right] \end{aligned} \quad (4.5)$$

donde

$$P'(c_i^L, t) = \sum_{c_j \supset c_i^L} P'(c_j, t) \quad \text{y} \quad P'(c_i^R, t) = \sum_{c_j \supset c_i^R} P'(c_j, t)$$

Obsérvese que el término de “reconstrucción” de la cuerda c_i depende de la aptitud de aquellas cuerdas que contienen los elementos c_i^R y c_i^L de c_i . Por lo tanto, el operador de entrecruzamiento introduce la noción de esquema. De esta forma se introducen los grados de libertad efectivos (esquemas) que se obtienen al sumar sobre todas las cuerdas que contienen a c_i^L o c_i^R . Estas cuerdas constituyen los grados de libertad microscópicos del sistema. Por otro lado, la ecuación de evolución 4.5 muestra que los efectos del término de reconstrucción son mayores que los del término destructivo si las partes de la cuerda son más selectivas que la cuerda completa.

A continuación se derivará la ecuación de evolución para esquemas que se obtiene al sumar la ecuación 4.5 sobre todas las cuerdas que contienen al esquema ξ . Primero, se considerará únicamente el operador de reproducción:

$$\text{Sea } P(\xi, t) = \sum_{c_i \supset \xi} P(c_i, t) \text{ entonces}$$

$$\begin{aligned} P(c_i, t+1) &= \frac{f_i P(c_i, t)}{\bar{f}(t)} \\ \Rightarrow \sum_{c_i \supset \xi} P(c_i, t+1) &= \frac{1}{\bar{f}(t)} \sum_{c_i \supset \xi} f_i P(c_i, t) \\ \Rightarrow P(\xi, t+1) &= \frac{\bar{f}(\xi, t) P(\xi, t)}{\bar{f}(t)} = P'(\xi, t) \end{aligned}$$

donde

$$\bar{f}(\xi, t) = \frac{\sum_{c_i \supset \xi} f(c_i, t) n(c_i, t)}{n(\xi, t)}$$

es la aptitud del esquema ξ .

Si se consideran los operadores de reproducción y mutación, la ecuación de evolución para esquemas es:

$$P(\xi, t+1) = \mathcal{P}(\xi) P'(\xi, t) + \sum_{\xi_i} \mathcal{P}(\xi_i \rightarrow \xi) P'(\xi_i, t)$$

donde ξ_i denota los esquemas que difieren de ξ en al menos un *bit* de los N_2 bits definidos en ξ . $P(\xi)$ es la probabilidad de que el esquema ξ no mute y $P(\xi_i \rightarrow \xi)$ es la probabilidad de que el esquema ξ_i mute al esquema ξ .

La ecuación de evolución para cuerdas considerando únicamente los operadores de selección y entrecruzamiento es:

$$\begin{aligned}
 P(c_i, t+1) &= P'(c_i, t) \\
 &\quad - \frac{p_c}{N-1} \sum_{k=1}^{N-1} [P'(c_i, t) - P'(c_i^L, t)P'(c_i^R, t)] \\
 \Rightarrow \sum_{c_i \supset \xi} P(c_i, t+1) &= \sum_{c_i \supset \xi} \left\{ P'(c_i, t) \right. \\
 &\quad \left. - \frac{p_c}{N-1} \sum_{k=1}^{N-1} [P'(c_i, t) - P'(c_i^L, t)P'(c_i^R, t)] \right\}
 \end{aligned}$$

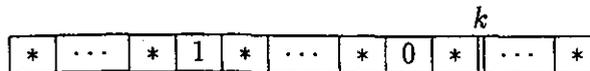
dado que $P'(c_i, t)$ no depende del punto de entrecruzamiento k , se tiene que

$$P(\xi, t+1) = (1 - p_c)P'(\xi, t) + \frac{p_c}{N-1} \sum_{k=1}^{N-1} \sum_{c_i \supset \xi} P'(c_i^L, t)P'(c_i^R, t)$$

Para simplificar esta ecuación, se divide la suma del término constructivo en aquellos términos donde el punto de entrecruzamiento k no corta el esquema ξ y en aquéllos donde sí lo corta:

$$\begin{aligned}
 P(\xi, t+1) &= (1 - p_c)P'(\xi, t) \\
 &\quad + \frac{p_c}{N-1} \sum_{c_i \supset \xi} \left[\sum_{k=1}^{N-l} P'(c_i^L, t)P'(c_i^R, t) + \sum_{k=1}^{l-1} P'(c_i^L, t)P'(c_i^R, t) \right]
 \end{aligned}$$

Se analizará el caso donde k no corta el esquema. Sin pérdida de generalidad, supóngase que k corta a la derecha del esquema. Por ejemplo, sea $\xi = 10$ un esquema de orden 2, entonces se tiene la siguiente situación:



Obsérvese que los *bits* en el esquema ξ a la derecha del punto de cruce k tienen un valor de *. Esto significa que c_i^R es la parte derecha del punto de cruce de cualquier cuerda. Entonces,

$$P'(c_i^R, t) = \sum_{c_i \supset c_i^R} P'(c_i, t) = 1$$

$$\Rightarrow \sum_{c_i \supset \xi} P'(c_i^L, t) P'(c_i^R, t) = \sum_{c_i \supset \xi} P'(c_i^L, t) = P'(\xi, t)$$

Por lo tanto,

$$P(\xi, t+1) = (1-p_c)P'(\xi, t) + \frac{p_c}{N-1}(N-l)P'(\xi, t)$$

$$+ \frac{p_c}{N-1} \sum_{c_i \supset \xi} \sum_{k=1}^{l-1} P'(c_i^L, t) P'(c_i^R, t)$$

$$= P'(\xi, t) - \frac{p_c}{N-1}(l-1)P'(\xi, t)$$

$$+ \frac{p_c}{N-1} \sum_{c_i \supset \xi} \sum_{k=1}^{l-1} P'(c_i^L, t) P'(c_i^R, t)$$

Supóngase que k corta el esquema ξ . El término $\sum_{c_i \supset \xi} P'(c_i^L, t) P'(c_i^R, t)$ es una suma sobre todos los posibles valores de los *bits* que no están definidos en el esquema ξ . Sean $n_L(n_R)$ los *bits* a la izquierda (derecha) del punto k que no están definidos en ξ . Entonces,

$$\sum_{c_i \supset \xi} P'(c_i^L, t) P'(c_i^R, t) = \sum_{n_L} \sum_{n_R} P'(c_i^L, t) P'(c_i^R, t) = P'(\xi_L, t) P'(\xi_R, t)$$

Por lo tanto,

$$P(\xi, t+1) = P'(\xi, t) - \frac{p_c}{N-1} \sum_{k=1}^{l-1} [P'(\xi, t) - P'(\xi_L, t) P'(\xi_R, t)] \quad (4.6)$$

Obsérvese que la suma es sobre los puntos de entrecruzamiento k que cortan al esquema ξ .

La interpretación de esta ecuación es similar al de la ecuación 4.5. En el término de reconstrucción, $P'(\xi_L, t) P'(\xi_R, t)$ es la probabilidad de seleccionar un padre que contenga la parte izquierda del esquema ξ y otro padre que contenga

la parte derecha de ξ . Además, un esquema se vuelve más numeroso por los efectos de entrecruzamiento si sus partes son más selectivas que el total.

La ecuación 4.6 es una extensión del teorema de esquemas. Obsérvese que la ecuación de evolución 4.6 toma en cuenta exactamente los efectos de "destrucción" y "reconstrucción" de cuerdas. Esta ecuación no es simplemente una cota inferior para la probabilidad de tener una cuerda C_i en la siguiente generación como lo establece el teorema de esquemas que únicamente toma en cuenta explícitamente el efecto "destrutivo" de entrecruzamiento.

La ecuación de evolución para esquemas que considera los efectos de reproducción, mutación y entrecruzamiento es:

$$P(\xi, t+1) = \mathcal{P}(\xi)P_c(\xi, t) + \sum_{\xi_i} \mathcal{P}(\xi_i \rightarrow \xi)P_c(\xi_i, t) \quad (4.7)$$

donde

$$P_c(\xi, t) = P'(\xi, t) - \frac{p_c}{N-1} \sum_{k=1}^{l-1} [P'(\xi, t) - P'(\xi_L, t)P'(\xi_R, t)]$$

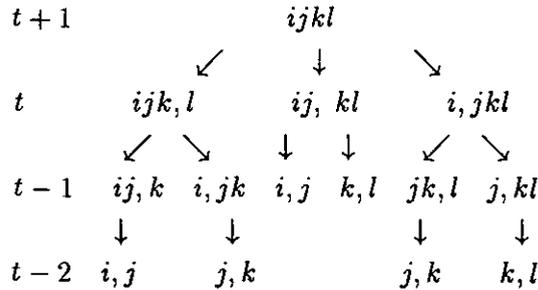
A continuación, se analizan algunas características de la ecuación 4.6 donde se incluyen los efectos de entrecruzamiento y se desprecia la mutación [7]. Obsérvese que para resolver la ecuación 4.6 se deben determinar primero los valores de $P(\xi_L, t)$ y $P(\xi_R, t)$. Pero $P(\xi_L, t)$ y $P(\xi_R, t)$ se obtienen con la ecuación de evolución, es decir

$$P'(\xi_L, t) = P'(\xi_L, t-1) - \frac{p_c}{N-1} \sum_{k=1}^{l-1} [P'(\xi_L, t-1) - P'(\xi_{LL}, t-1)P'(\xi_{LR}, t-1)]$$

donde ξ_{LL} y ξ_{LR} son la parte izquierda y derecha del esquema ξ_L con respecto al punto de entrecruzamiento m donde $m < k$. Además, ξ_{LL} y ξ_{LR} son esquemas de menor orden que ξ_L . Este proceso continúa. Para calcular $P(\xi, t+1)$ se necesitan $P(\xi_L, t)$ y $P(\xi_R, t)$ que a su vez requieren de $P(\xi_{LL}, t-1)$, $P(\xi_{LR}, t-1)$, $P(\xi_{RL}, t-1)$ y $P(\xi_{RR}, t-1)$, etc. En cada paso, los esquemas disminuyen de orden. El proceso termina con los esquemas de orden 1, los cuales no se destruyen con el operador de entrecruzamiento. Es decir, para esquemas de orden 1, $\xi = i$, la ecuación 4.6 se simplifica como:

$$P(i, t+1) = P'(i, t)$$

Como ejemplo, considérese un esquema de orden 4, $\xi = ijkl$. La estructura jerárquica de los esquemas que se necesitan para formar ξ es la siguiente:



En este ejemplo, se ilustra claramente la idea de bloques constructores. Los 4 bloques constructores de orden 1 (i, j, k , y l) se combinan para formar bloques constructores de orden 2 (ij, kl y jk) que a su vez se combinan con bloques constructores de orden 1 para formar bloques constructores de orden 3 (ijk y jkl), etc.

A continuación se analizan las características de los grados de libertad efectivos en un paisaje de aptitud plano, es decir, donde todas las cuerdas tienen la misma aptitud. Sin pérdida de generalidad, considérese la ecuación de evolución para el esquema $\xi = 11$ de orden 2:

$$P(11, t + 1) = P'(11, t) - \frac{p_c}{N - 1} \sum_{k=1}^{l-1} [P'(11, t) - P'(1_L, t)P'(1_R, t)]$$

donde 1_L y 1_R son esquemas de orden 1 que forman la parte izquierda y derecha de $\xi = 11$ respecto al punto de entrecruzamiento k .

$$\Rightarrow P(11, t + 1) = P'(11, t) - p_c \frac{(l - 1)}{(N - 1)} [P'(11, t) - P'(1_L, t)P'(1_R, t)]$$

Se sabe que $P'(11, t) = \frac{\bar{f}(11, t)}{\bar{f}(t)} P(11, t)$, entonces

$$\begin{aligned}
 P(11, t + 1) &= \frac{\bar{f}(11, t)}{\bar{f}(t)} \left[1 - p_c \frac{(l - 1)}{(N - 1)} \right] P(11, t) \\
 &\quad + p_c \frac{(l - 1)}{(N - 1)} P'(1_L, t)P'(1_R, t)
 \end{aligned}$$

Se desea saber a qué converge $P(11)$. Para ello, se obtiene la ecuación diferencial $\frac{dP(11,t)}{dt}$:

$$\begin{aligned}\frac{dP(11,t)}{dt} &= \lim_{\delta t \rightarrow 0} \frac{P(11,t+\delta t) - P(11,t)}{\delta t} \\ &= -\frac{\bar{f}(11,t)}{f(t)} p_c \frac{(l-1)}{(N-1)} P(11,t) \\ &\quad + p_c \frac{(l-1)}{(N-1)} P'(1_L,t) P'(1_R,t)\end{aligned}$$

Obsérvese que $P(1,t+1) = P'(1,t) = \frac{\bar{f}(1,t)P(1,t)}{f(t)}$, pero en un paisaje de aptitud plano $\frac{\bar{f}(1,t)}{f(t)} = 1$ entonces $P(1,t+1) = P(1,t)$. Por lo tanto

$$\frac{dP(11,t)}{dt} = -p_c \frac{(l-1)}{(N-1)} [P(11,t) - P(1_L,0)P(1_R,0)]$$

Al resolver esta ecuación diferencial de primer orden se obtiene

$$P(11,t) = P(11,0)e^{-p_c \left(\frac{l-1}{N-1}\right)t} + P(1_L,0)P(1_R,0) \left[1 - e^{-p_c \left(\frac{l-1}{N-1}\right)t}\right]$$

Si $t \rightarrow \infty$, $P(11,t) = P(1_L,0)P(1_R,0)$. Esto significa que en el punto estable (es decir, en el límite cuando la población no se modifica) aparece el esquema $\xi = 11$ debido al efecto "constructivo" de entrecruzamiento. Obsérvese que sin este efecto constructivo, $P(11)$ convergería a cero. Por lo tanto, el término que domina es el constructivo y no el destructivo como se plantea en la Hipótesis de Bloques Constructores. Esto significa que los esquemas de longitud grande no se destruyen. Por lo tanto, los grados de libertad efectivos de un AG no son necesariamente esquemas de longitud pequeña. La idea de bloques constructores está asociada a los *bits* definidos de los esquemas que construyen soluciones aptas. Pero los *bits* definidos no necesariamente tienen que estar juntos en la cuerda. El resultado que se acaba de mostrar es válido sólo en paisajes de aptitud neutrales donde no hay preferencia por esquemas de cierta longitud. Es decir, donde los esquemas de determinada longitud no sean más aptos que los demás. Cabe hacer notar que la Hipótesis de Bloques Constructores trata de describir las características de los grados de libertad efectivos sin tomar en cuenta el paisaje de aptitud bajo consideración, lo cual es una afirmación muy fuerte.

4.3 Aptitud Efectiva

Otra medida que proporciona información de la evolución de esquemas es la **aptitud efectiva** [3]. La idea intuitiva de aptitud es que los padres más aptos tienen más descendientes. Sin embargo, se analizará la noción de aptitud con el siguiente ejemplo. Se considerará el efecto de entrecruzamiento sin mutación en un modelo que consiste de esquemas de orden 2: 11, 10, 01 y 00. El paisaje de aptitud es: $f(11) = f(00) = 1$ y $f(10) = f(01) = 0$. A partir de la ecuación 4.6 se obtienen las ecuaciones de evolución para esquemas de orden 2. Por ejemplo, para $\xi = 11$ la ecuación 4.6 se simplifica de la siguiente manera:

$$\begin{aligned}
 P(11, t+1) &= \frac{\bar{f}(11, t)}{\bar{f}(t)} P(11, t) \\
 &\quad - p_c \frac{(l-1)}{(N-1)} \left[\frac{\bar{f}(11, t)}{\bar{f}(t)} P(11, t) \frac{\bar{f}(00, t)}{\bar{f}(t)} P(00, t) \right. \\
 &\quad \left. - \frac{\bar{f}(01, t)}{\bar{f}(t)} P(01, t) \frac{\bar{f}(10, t)}{\bar{f}(t)} P(10, t) \right]
 \end{aligned}$$

La solución de $P(\xi, t)$ en el estado estacionario, es decir cuando $P(\xi, t+1) = P(\xi, t)$ es la siguiente:

$$\begin{aligned}
 P(11) = P(00) &= \frac{1}{2} \left[1 - \frac{p_c(l-1)}{2(N-1)} \right] \\
 P(01) = P(10) &= \frac{p_c(l-1)}{4(N-1)}
 \end{aligned}$$

Para $l = N$ y $p_c = 1$ se tiene que $P(11) = P(00) = P(10) = P(01) = \frac{1}{4}$. Es decir, la mitad de la población está formada por cuerdas que tienen aptitud cero. Pero obsérvese que estas cuerdas no desaparecen porque al cruzarlas entre ellas se forman cuerdas más aptas.

Este ejemplo muestra cómo los operadores genéticos pueden cambiar el paisaje de aptitud "efectivo" en el que evoluciona la población. En este caso, el paisaje de aptitud no proporciona suficiente información acerca de cómo evoluciona la población, ya que una cuerda con aptitud cero puede estar presente en las siguientes generaciones. Por lo tanto, se puede definir la aptitud efectiva $f_{eff}(\xi, t)$ de la siguiente manera:

$$P(\xi, t+1) = \frac{f_{eff}(\xi, t)}{\bar{f}(t)} P(\xi, t)$$

Donde $P(\xi, t + 1)$ satisface la ecuación 4.7. Si se toman en cuenta únicamente los operadores de selección y entrecruzamiento se obtiene la siguiente expresión para $f_{eff}(\xi, t)$:

$$f_{eff}(\xi, t) = \bar{f}(\xi, t) - \frac{\bar{f}(t)p_c}{P(\xi, t)(N-1)} \sum_{k=1}^{l-1} [P'(\xi, t) - P'(\xi_L, t)P'(\xi_R, t)]$$

Obsérvese que si $p_c = 0$ entonces $f_{eff}(\xi, t) = \bar{f}(\xi, t)$.

CAPÍTULO 5

Resultados Experimentales de Grados de Libertad Efectivos.

En este capítulo, se presentan resultados experimentales de los efectos que tiene entrecruzamiento en la longitud del esquema y se comparan con los resultados teóricos obtenidos en el capítulo 4. Para ello, se implementó un *AG* sin mutación ya que éste operador no tiene ningún efecto directo sobre la longitud del esquema.

Las características de los grados de libertad efectivos dependen del paisaje de aptitud bajo consideración. Sin embargo, para observar los efectos genéricos de entrecruzamiento se debe elegir un paisaje de aptitud neutral en ausencia de este operador. Es decir, un paisaje de aptitud donde la selección no favorezca esquemas de determinada longitud. El modelo descrito en el capítulo 2, donde la aptitud de una cuerda C_i es igual al número de unos que tiene, cumple con estas características. En este modelo $f(C_i) = \sum_j 1_j$ donde 1_j es 1 (0) si el j -ésimo *bit* de la cuerda vale 1 (0).

Los grados de libertad efectivos son esquemas asociados con el óptimo que en este caso es $C^* = 11 \cdots 1$. Sin pérdida de generalidad, se eligieron esquemas de orden 2 asociados con el óptimo, es decir, $\xi = 11$ y cuerdas de tamaño $N = 8$. Se desea determinar el número de esquemas $\xi = 11$ de longitud $l = 2, 3, \dots, 8$ en cada generación ($n'_{opt}(l)$). Obsérvese que un esquema $\xi = 11$ de longitud 7 puede ubicarse en 2 posiciones diferentes en una cuerda de tamaño 8, $1****1*$ y $*1****1$. En cambio, un esquema $\xi = 11$ de longitud 8 sólo puede encontrarse en una posición dentro de la cuerda, a saber en $1****1$. Entonces, para poder comparar el número de esquemas de orden 2 de diferente longitud que hay en una población, se debe normalizar ($n'_{opt}(l)$) con el número de posiciones diferentes que puede ocupar el esquema en una cuerda de tamaño 8:

$$n_{opt}(l) = \frac{n'_{opt}(l)}{8 - l + 1} \quad l = 2, \dots, 8$$

Con el objeto de evitar los efectos de tamaño finito, se eligió una población de

5000 individuos. Las gráficas corresponden al promedio de 30 corridas diferentes. Las gráficas 5.1 y 5.2 muestran $M(l)$ como función del tiempo, donde

$$M(l) = \frac{n_{opt}(l) - n_{opt}(8)}{n_{opt}(8)} \quad l = 2, \dots, 7$$

En la gráfica 5.1 no se utilizó entrecruzamiento. Obsérvese que no hay preferencia por esquemas de alguna longitud. La gráfica 5.2 muestra la evolución de esquemas cuando $p_c = 1.0$. En este caso, los esquemas están ordenados de acuerdo a su longitud de tal manera que los esquemas de mayor longitud son los más numerosos. Entonces, para este modelo la reconstrucción de esquemas es más importante que la destrucción de éstos. Estos resultados concuerdan con los obtenidos teóricamente a partir de las ecuaciones de evolución.

Las gráficas 5.3 y 5.4 muestran la evolución de la aptitud efectiva de los esquemas de determinada longitud correspondientes a las gráficas 5.1 y 5.2. Es decir, se grafica $F(l)$ como función del tiempo, donde

$$F(l) = \frac{f_{eff}(l) - f_{eff}(8)}{f_{eff}(8)}$$

Sea $f_{eff}^j(l)$ la aptitud efectiva del esquema $\xi = 11$ de longitud l cuyo primer *bit* definido se encuentra en la posición j de la cadena. Entonces,

$$f_{eff}(l) = \frac{\sum_j f_{eff}^j(l)}{8 - l + 1} \quad \text{con } l = 2, \dots, 8 \text{ y } j = 1, \dots, 8 - l + 1$$

En el apéndice, se incluye el código de las subrutinas que calculan $f_{eff}(l)$ y $n_{opt}(l)$. La aptitud efectiva es otra manera de medir el efecto de entrecruzamiento ya que entre mayor sea la aptitud efectiva de un esquema, más numeroso será éste. La gráfica 5.4 corresponde al caso en el que $p_c = 1.0$. Obsérvese que los esquemas de longitud grande tienen una aptitud efectiva mayor o igual que los de longitud pequeña en las primeras 5 generaciones aproximadamente. La ventaja en la aptitud efectiva del esquema $\xi = 11$ de longitud 8 sobre el de longitud 2 es de 0.01. En cambio, sin entrecruzamiento (gráfica 5.3) las fluctuaciones son del orden de 0.0005. Por lo tanto, con entrecruzamiento el esquema $\xi = 11$ de longitud 8 tiene una ventaja selectiva 20 veces mayor aproximadamente sobre el esquema $\xi = 11$ de longitud 2.

Dado que no hay mutaciones, la población en la distribución estacionaria está formada únicamente por el óptimo 11111111. Por lo tanto, el número de esquemas $\xi = 11$ de longitud $l = 2, 3, \dots, 8$ debe ser igual al tamaño de la población que en este caso es 5000. Es decir,

$$\frac{n_{opt}(l)}{8-l+1} = 5000 \quad \text{para } l = 2, 3, \dots, 8$$

Pero en las primeras generaciones, los esquemas de longitud pequeña son menos numerosos que los de longitud grande. Entonces, los esquemas de longitud pequeña deben incrementarse más que los de longitud grande para alcanzar una concentración de 5000 en la población. Esto se refleja en la gráfica 5.4 donde los esquemas de longitud pequeña tienen una aptitud efectiva mayor a los de longitud grande a partir de la sexta generación aproximadamente.

Una vez que se han estudiado las propiedades genéricas de entrecruzamiento, se pueden considerar paisajes de aptitud más complicados para observar los efectos de selección y entrecruzamiento conjuntamente. En el siguiente paisaje de aptitud, la selección introduce un sesgo hacia los esquemas de determinada longitud. La aptitud de la cuerda c_i es:

$$f(c_i) = \sum_j 1_j + \frac{\epsilon}{N^\pm} \sum_{j < k} l_{jk}^{\pm 1}$$

donde l_{jk} es la longitud del esquema $\xi = 11$ cuyos unos están ubicados en las posiciones j y k de la cuerda c_i . Para l_{jk} se tiene un sesgo hacia los esquemas de longitud grande ya que éstos tienen una aptitud mayor. Por otro lado, l_{jk}^{-1} introduce un sesgo hacia los esquemas de longitud menor. $N^+ = \sum_{j < k} l_{jk}$ ($N^- = \sum_{j < k} l_{jk}^{-1}$) es un factor de normalización que consiste en la suma sobre todas las longitudes (inverso de las longitudes) de los esquemas $\xi = 11$ asociados con el óptimo 11111111. El parámetro ϵ controla la magnitud del sesgo. Entonces, la aptitud de la cuerda $c_i = 11001100$ es:

$$f(c_i) = 4 + \frac{\epsilon}{112} [(2 + 5 + 6) + (4 + 5) + 2]$$

si se favorece a los esquemas de longitud grande y

$$f(c_i) = 4 + \epsilon \left(\frac{280}{2369} \right) \left[\left(\frac{1}{2} + \frac{1}{5} + \frac{1}{6} \right) + \left(\frac{1}{4} + \frac{1}{5} \right) + \frac{1}{2} \right]$$

si el sesgo es para los esquemas de longitud pequeña.

La gráfica 5.5 muestra la evolución de esquemas para el modelo donde los esquemas de longitud grande tienen preferencia. Se fijó $\epsilon = 0.3$ y $p_c = 0.0$. Obsérvese que esta gráfica es similar a la gráfica 5.2. Es decir, el efecto que tiene

entrecruzamiento es el de favorecer a los esquemas de longitud grande. La gráfica 5.6 representa la evolución de esquemas para el caso en el que los esquemas de longitud pequeña tienen preferencia. En este caso, $p_c = 1.0$ y $\epsilon = 0.75$. Obsérvese que entrecruzamiento anula el sesgo inducido por selección. Las gráficas 5.7 y 5.8 muestran la evolución de la aptitud efectiva de los esquemas correspondientes a las gráficas 5.5 y 5.6 respectivamente.

Por último, se consideró un paisaje de aptitud deceptivo. En un problema deceptivo, los esquemas aptos de orden bajo construyen esquemas de orden mayor pero con aptitud baja. Es decir, los esquemas de orden bajo asociados con subóptimos tienen una aptitud mayor que aquellos esquemas de orden bajo asociados con el óptimo. En este caso, el paisaje de aptitud se definió de la siguiente manera. Cada pareja de *bits* contribuye de la siguiente manera en la aptitud de la cuerda: $f(11) = 3$, $f(01) = f(10) = 1$ y $f(00) = 2$. La aptitud de una cuerda c_i se define como

$$f(c_i) = \frac{1}{nC_2} \sum_{j < k} f_{jk}$$

Por ejemplo, si $c_i = 010$,

$$f(c_i) = \frac{1}{3C_2} (f(01) + f(00) + f(10)) = \frac{1}{3}(1 + 2 + 1)$$

Las gráficas 5.9 y 5.10 representan la evolución de esquemas cuando $p_c = 0.0$ y $p_c = 1.0$ respectivamente. Obsérvese que en este modelo, entrecruzamiento favorece los esquemas de longitud pequeña ya que el efecto destructivo de entrecruzamiento es más importante que el constructivo. Los esquemas $\xi = 10$ y $\xi = 01$ tienen una concentración baja en la población porque el problema es deceptivo. Por lo tanto, es poco probable construir el esquema $\xi = 11$ a partir de $\xi = 10$ y de $\xi = 01$. Por otro lado, es claro que los esquemas $\xi = 11$ de longitud grande se rompen más fácilmente que los de longitud pequeña debido al efecto destructivo de entrecruzamiento. Entonces, los esquemas $\xi = 11$ que se ven favorecidos en este caso son los de longitud pequeña.

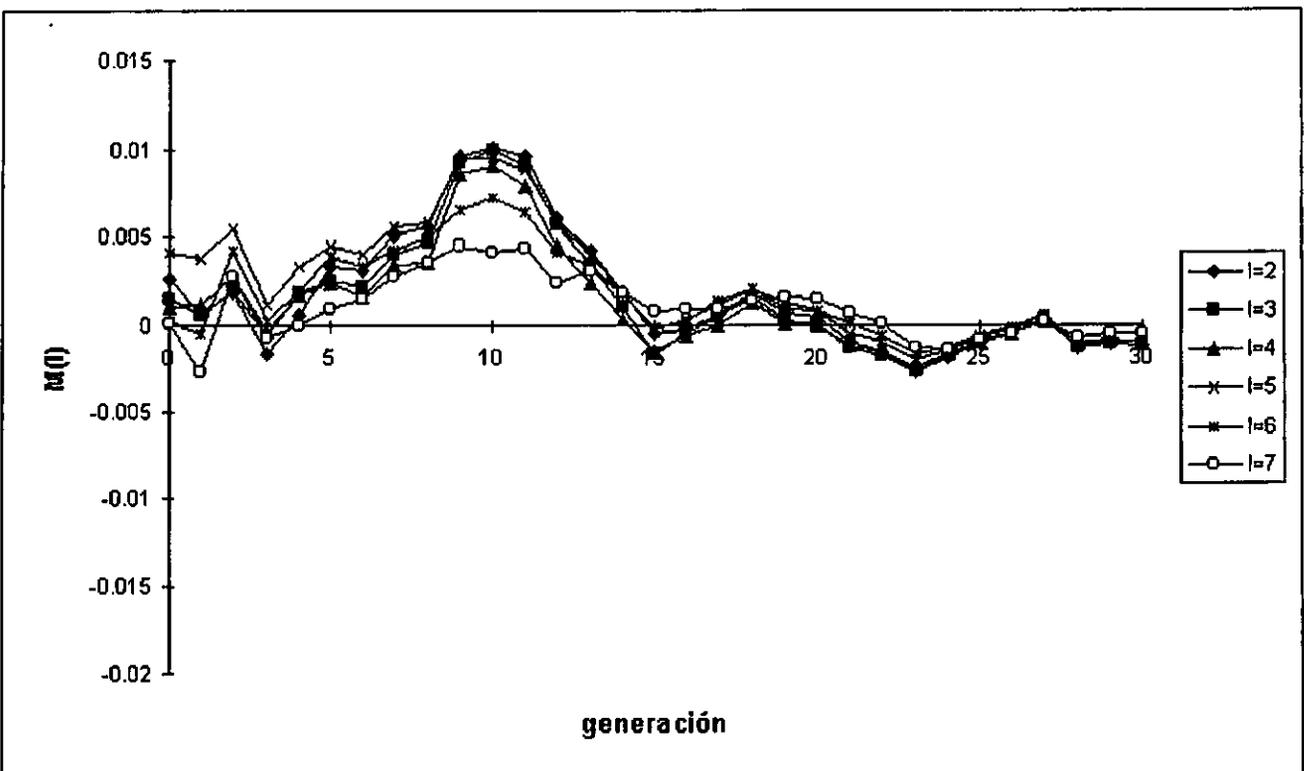


Figura 5.1: Gráfica de $M(t)$ como función de t para el modelo donde se cuenta el número de unos con $p_c = 0.0$.

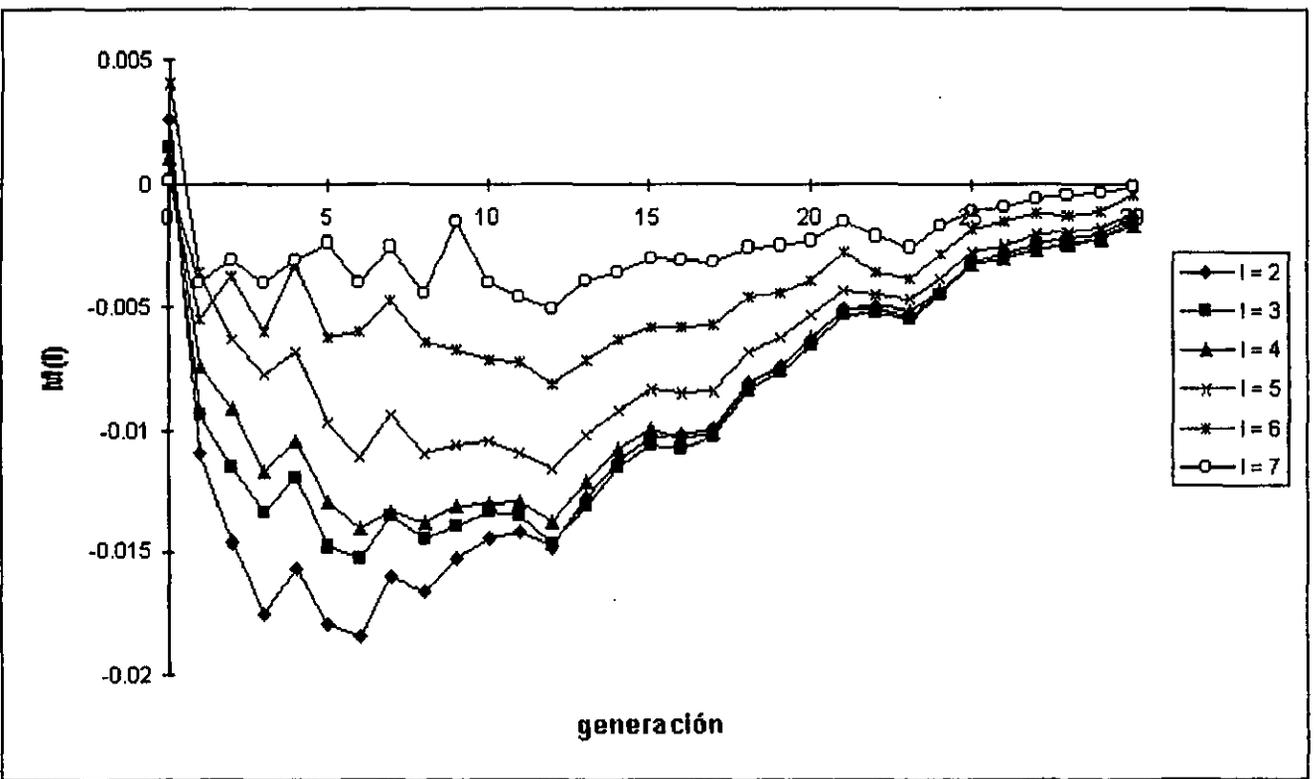


Figura 5.2: Gráfica de $M(t)$ como función de t para el modelo donde se cuenta el número de unos con $p_c = 1.0$.

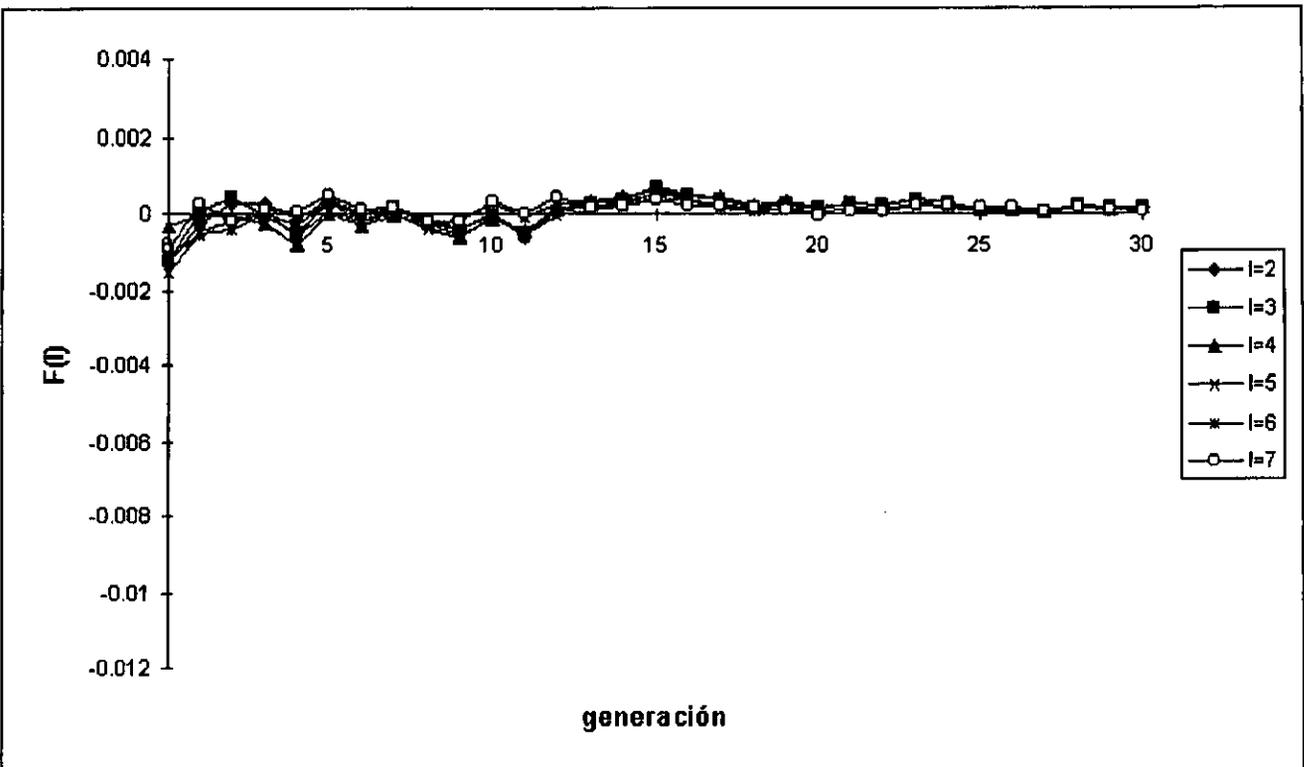


Figura 5.3: Gráfica de $F(t)$ como función de t para el modelo donde se cuenta el número de unos con $p_c = 0.0$.

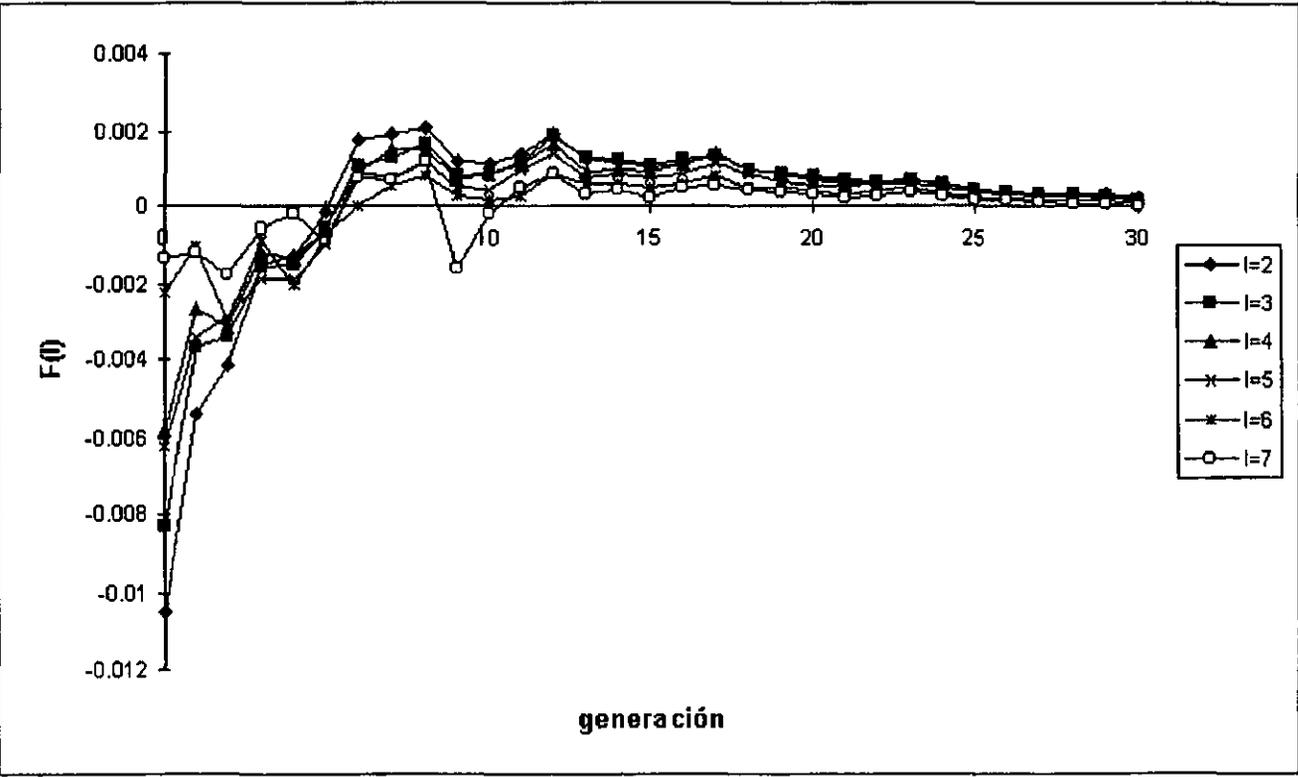


Figura 5.4: Gráfica de $F(t)$ como función de t para el modelo donde se cuenta el número de unos con $p_e = 1.0$.

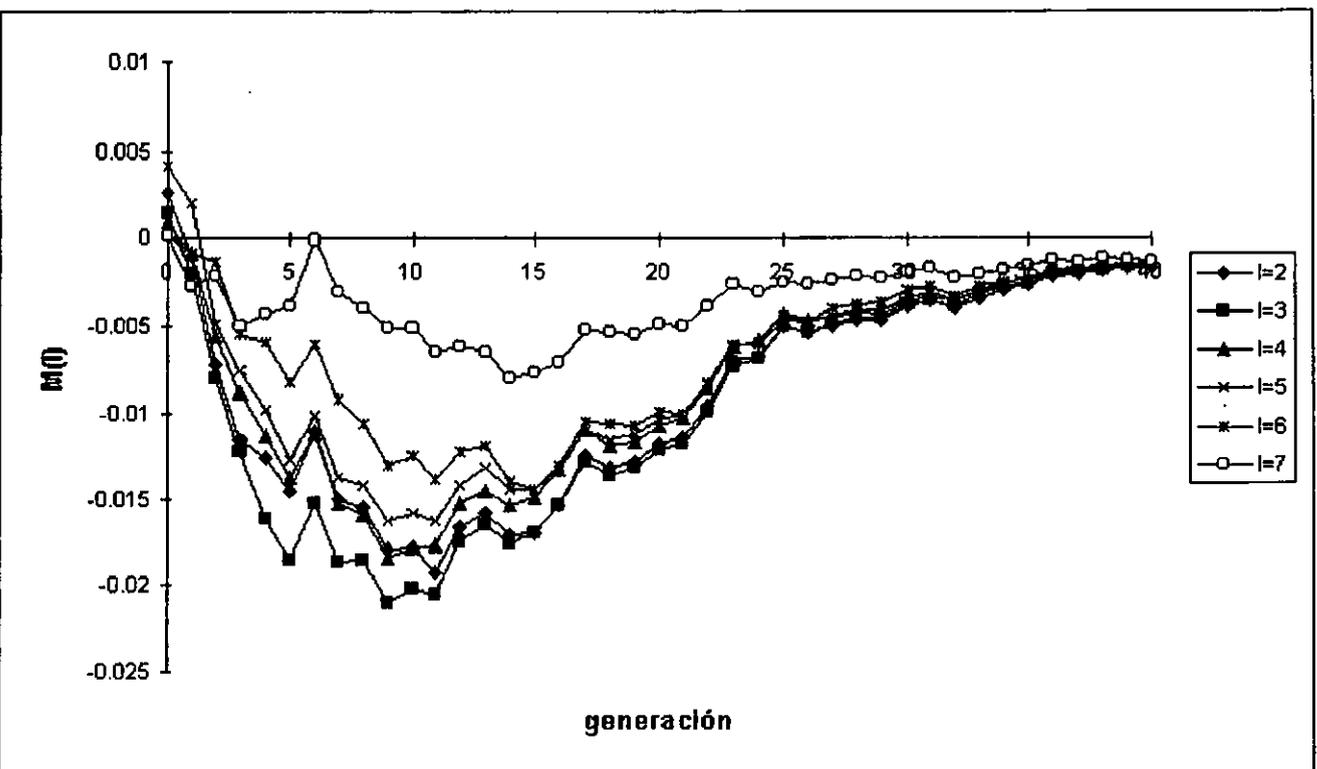


Figura 5.5: Gráfica de $M(t)$ como función de t para el modelo donde los esquemas de longitud grande tienen preferencia con $\epsilon = 0.3$ y $p_c = 0.0$.

SALIR ESTA TESIS NO DEBE
 DE LA BIBLIOTECA

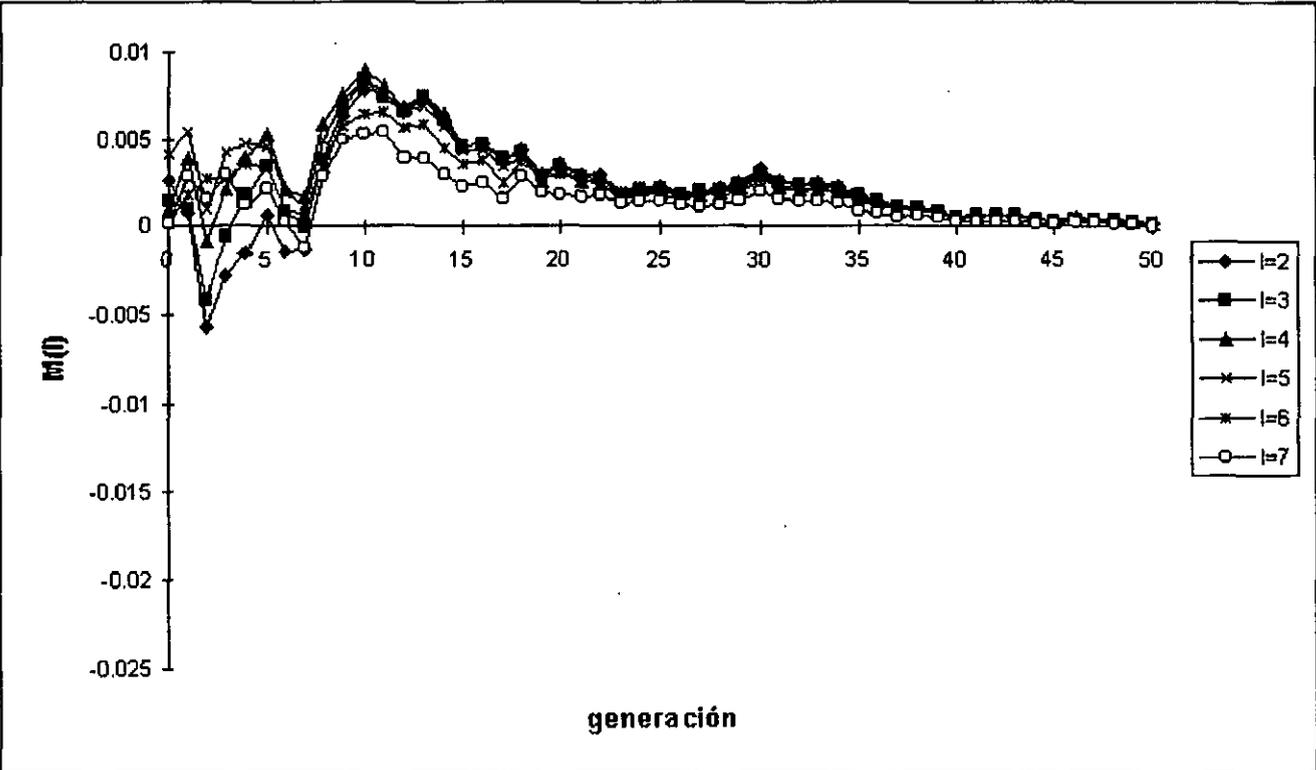


Figura 5.6: Gráfica de $M(t)$ como función de t para el modelo donde los esquemas de longitud pequeña tienen preferencia con $\epsilon = 0.75$ y $p_c = 1.0$.

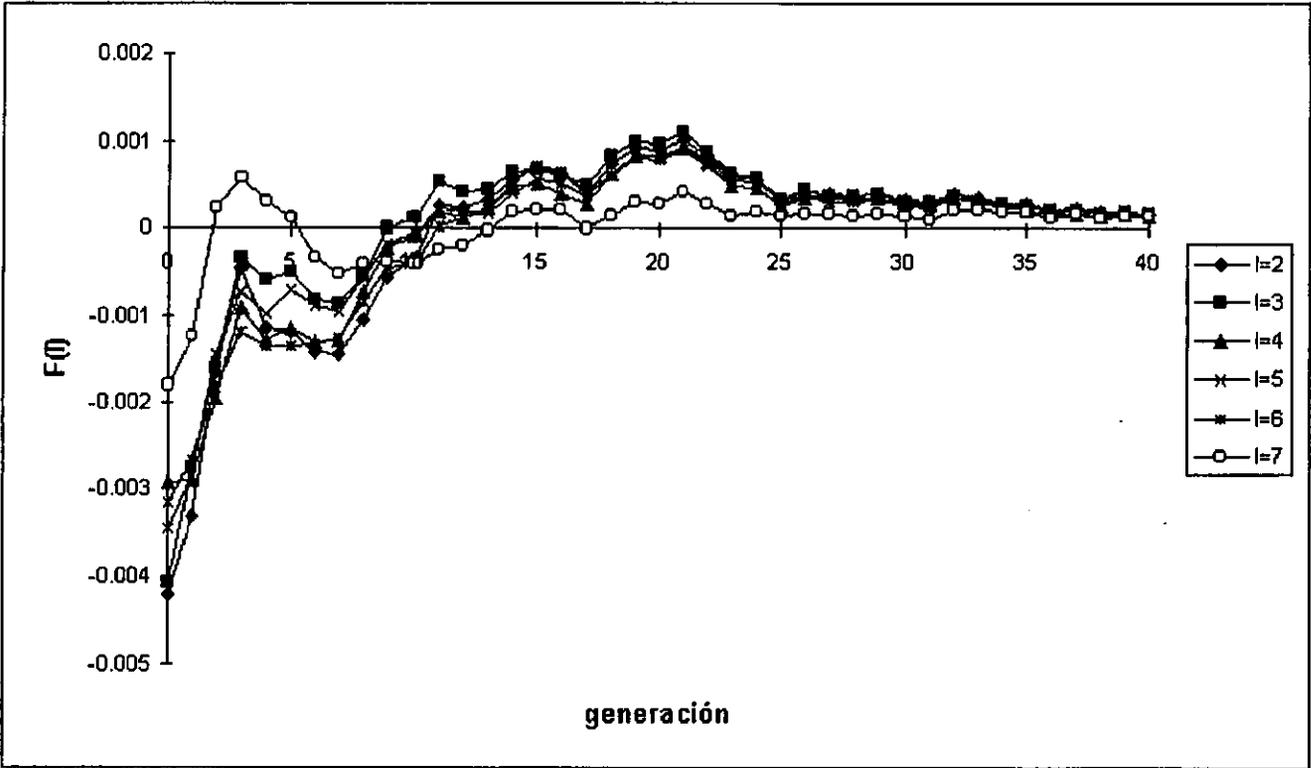


Figura 5.7: Gráfica de $F(t)$ como función de t para el modelo con sesgo de la figura 5.5.

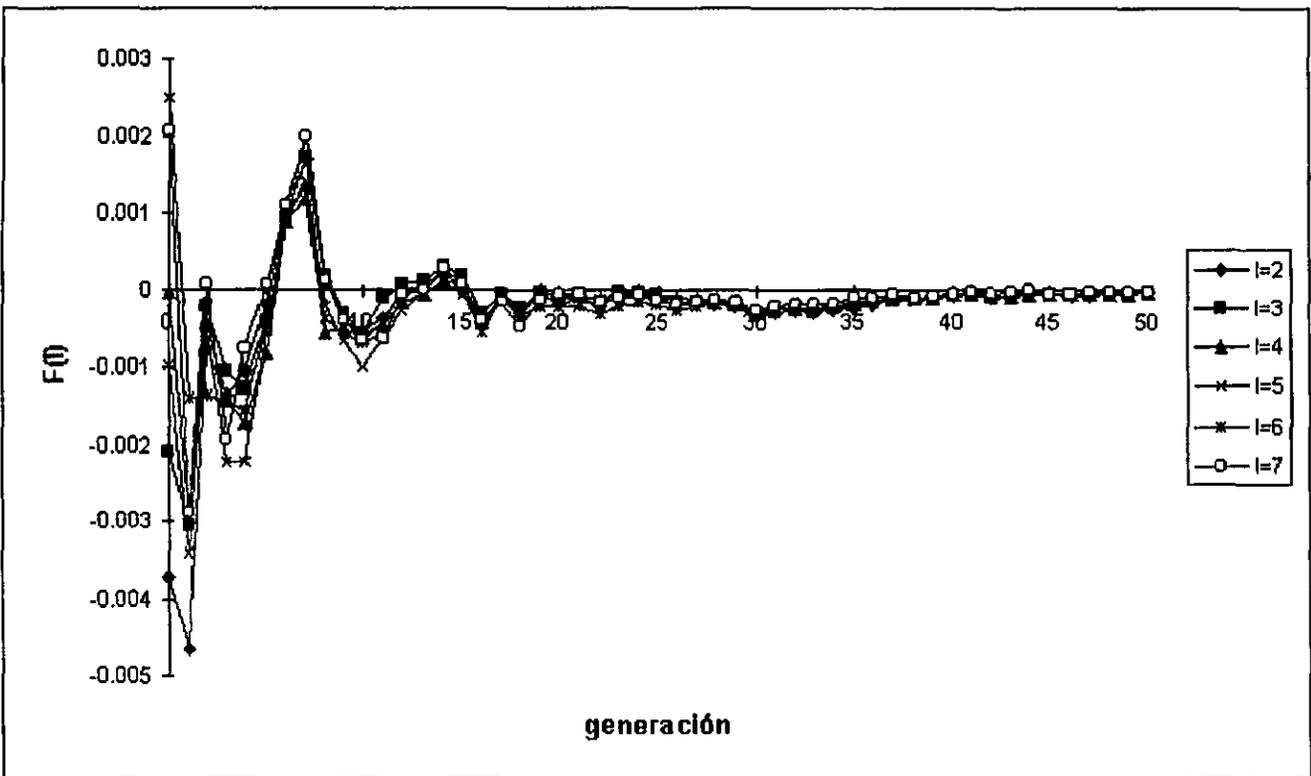


Figura 5.8: Gráfica de $F(t)$ como función de t para el modelo con sesgo de la figura 5.6.

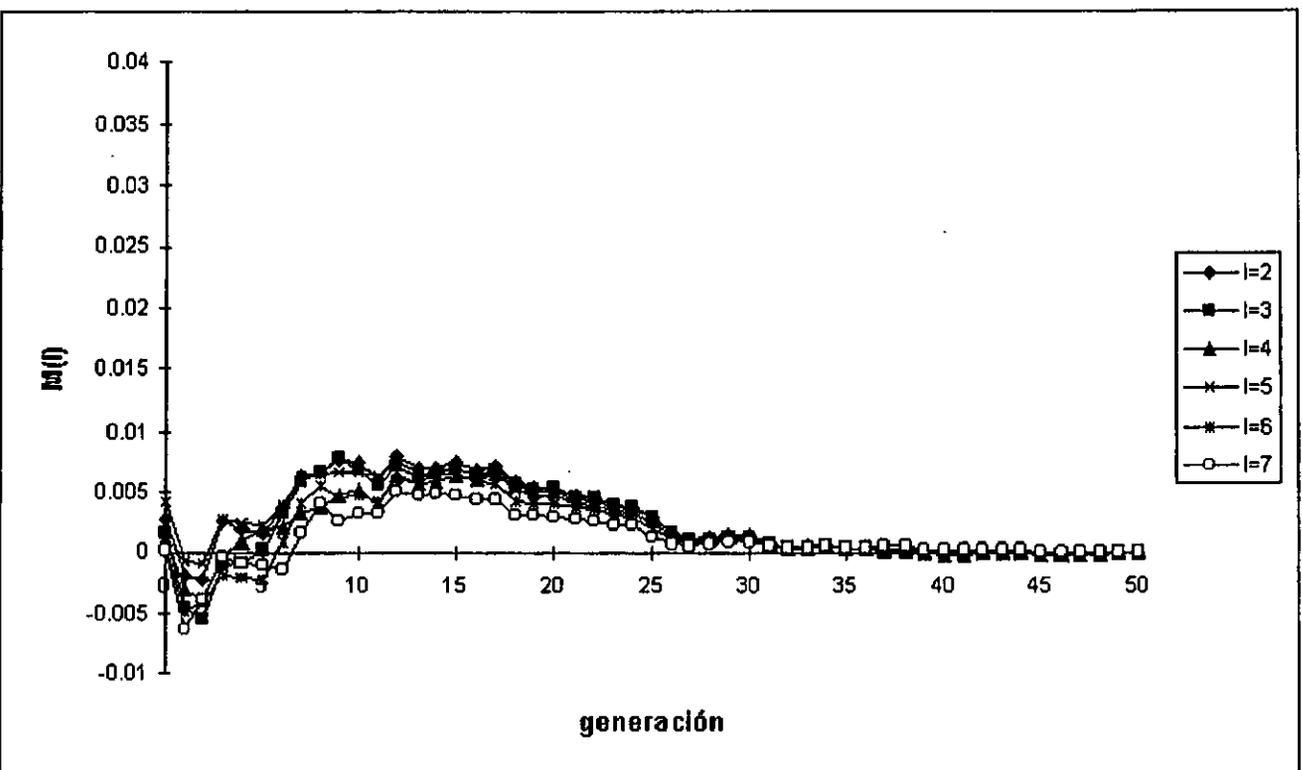


Figura 5.9: Gráfica de $M(t)$ como función de t para el modelo deceptivo con $p_c = 0.0$.

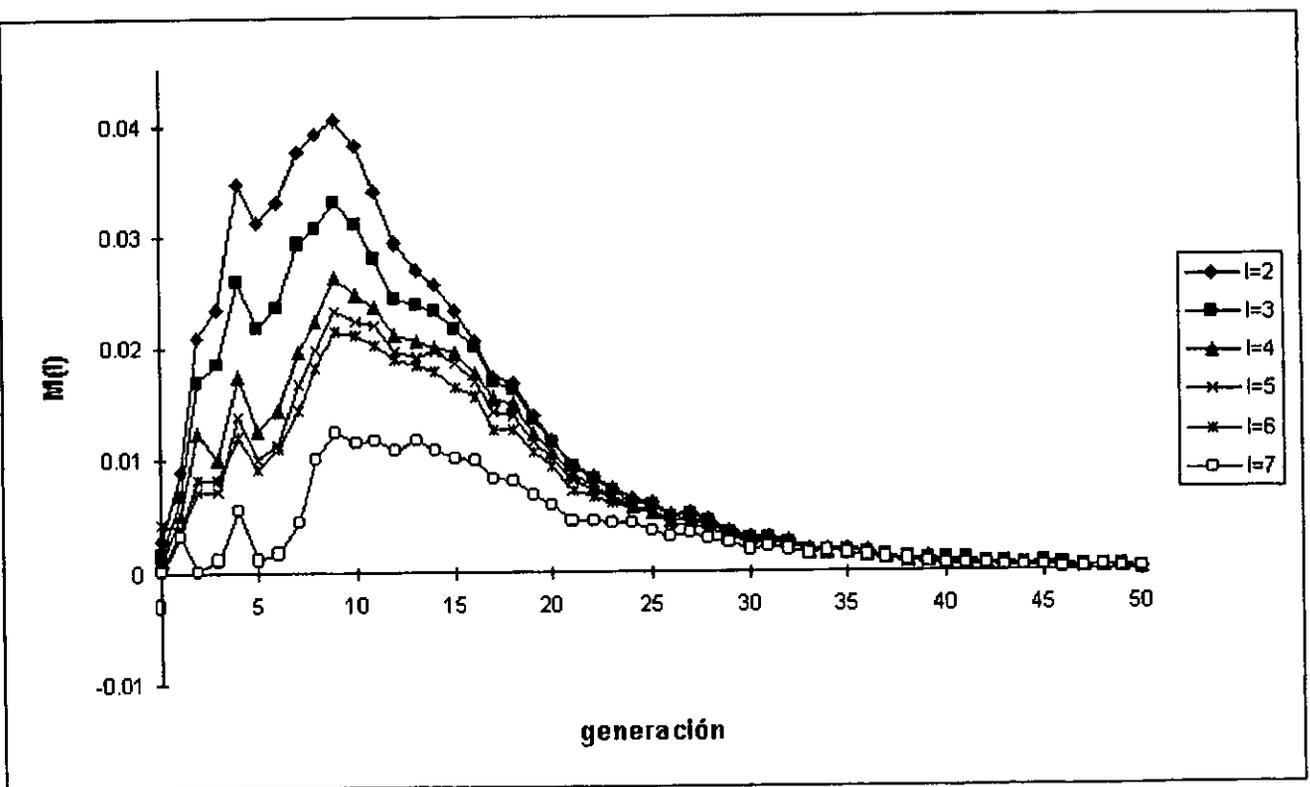


Figura 5.10: Gráfica de $M(t)$ como función de t para el modelo deceptivo con $p_c = 1.0$.

CONCLUSIONES

En este trabajo, se estudiaron los grados de libertad efectivos de los AGs con el objeto de entender cómo busca un AG las soluciones aptas de un problema. Esto permite implementar de manera más adecuada los AGs para resolver determinados problemas. Sin embargo, también permite estudiar los fenómenos de la genética de poblaciones y la selección natural ya que los AGs son modelos de la naturaleza. Los resultados presentados en este trabajo se respaldaron con resultados experimentales y además se compararon con la teoría estándar de AGs.

Para estudiar los grados de libertad efectivos de los AGs, se presentó una ecuación de evolución exacta a partir de la cual se obtiene la distribución de probabilidad de esquemas como función del tiempo cuando la población es infinita. Esta ecuación proporciona más información que la cota inferior determinada en el teorema de esquemas ya que se considera el efecto constructivo del operador de entrecruzamiento. Se mostró que este operador introduce la noción de esquema y que los esquemas aptos, de orden bajo y longitud grande (grados de libertad efectivos) construyen esquemas de orden mayor cuando el efecto de reconstrucción de esquemas es más importante que el de destrucción. Es decir, se demostró que entrecruzamiento favorece los esquemas de longitud grande en paisajes de aptitud neutrales contrariamente a lo que se afirma en la Hipótesis de Bloques Constructores. Para comprobar experimentalmente estos resultados, se utilizaron modelos simples y se eliminaron las mutaciones con el objeto de observar únicamente los efectos de entrecruzamiento. En el paisaje de aptitud neutral (donde se cuenta el número de unos) la concentración de los esquemas en la población es inversamente proporcional a su longitud. Para el paisaje de aptitud donde se introduce un sesgo a los esquemas de determinada longitud, se observó que el efecto de entrecruzamiento es análogo al efecto que se tiene cuando se favorecen esquemas de longitud grande. Por otro lado, en el problema deceptivo, la concentración de los esquemas pequeños en la población es mayor porque el efecto destructivo de entrecruzamiento es más importante en este caso. Estos resultados concuerdan con los obtenidos teóricamente. Sin embargo, es conveniente considerar más paisajes de aptitud. Los modelos NK -Kauffman [8] proporcionan una gran diversidad de paisajes de aptitud ya que se puede variar el tamaño de la cuerda (N) y el número de interacciones epistáticas entre

los *bits* de ésta (K). También se puede hacer un estudio más detallado de los caminos elegidos por el *AG* en el problema deceptivo. Dado que hay preferencia por cuerdas de aptitud alta, las rutas no deceptivas son las que posiblemente tienen mayor preferencia.

Por otro lado, también se estudió el comportamiento de los grados de libertad efectivos cuando se modifican los parámetros. Se observó que un *AG* presenta una fase ordenada dominada por selección (cuando p es pequeño) y otra desordenada dominada por mutación (cuando p es grande). La tasa de mutación crítica que divide estos 2 comportamientos es el umbral de error. A este fenómeno se le conoce como transición de fase el cual se ha estudiado más a fondo en la física. Por esta razón se aprovecharon los resultados de la mecánica estadística para estudiar los efectos de los operadores genéticos bajo un marco formal. Además, se obtuvo experimentalmente el umbral de error en el paisaje de aptitud de la secuencia maestra y en el paisaje de aptitud plano con 2 picos estudiados por Tarazona. El primer paisaje de aptitud presenta una transición de fase de segundo orden y el segundo, una transición de fase de primer orden. Estos resultados coinciden con los obtenidos teóricamente por Tarazona con técnicas de la mecánica estadística. También se variaron algunos parámetros del *AG* en el paisaje de aptitud de la secuencia maestra para observar cómo se modifica el umbral de error. Se encontró que entre menor sea la presión selectiva del óptimo, menor es el umbral de error. Por otro lado, se observó que el umbral de error es inversamente proporcional al tamaño de la cuerda. Estas observaciones permiten fijar el valor de la tasa de mutación en un *AG* de manera más adecuada. De tal forma que la tasa de mutación debe disminuirse si el óptimo no tiene mucha ventaja selectiva sobre las demás cuerdas o si el tamaño de las cuerdas es grande. También es conveniente trabajar con poblaciones grandes, con respecto al número de estados posibles, para evitar los efectos de tamaño finito de la población. Sin embargo, sería interesante determinar el umbral de error en paisajes de aptitud más complicados para observar si se tiene el mismo comportamiento. También se puede incluir el operador de entrecruzamiento para estudiar los efectos que tiene sobre el umbral de error [9].

APÉNDICE A

Subrutinas del Programa del AG

En este apéndice se incluye el código, escrito en C, de las subrutinas que calculan $n_{opt}(l)$ y $f_{eff}(l)$. Es decir, el número normalizado de esquemas $\xi = 11$ de longitud $l = 2, 3, \dots, 8$ que hay en la población y su respectiva aptitud efectiva. Estas subrutinas se incluyeron en el programa "SGA-C: A C-Language Implementation of a Simple Genetic Algorithm" de Goldberg, David E. Smith y Jeff A. Erickson [2].

```

/* Estructuras y variables globales. */
struct individual
{
    unsigned *chrom;
        /* cromosoma o cadena de un individuo */
    double fitness;
        /* aptitud de un individuo */
}

double esq[7];
    /* Arreglo con el número de esquemas  $\xi = 11$  de longitud  $l = 2, 3, \dots, 8$  */
double f_eff[7];
    /* Arreglo con la aptitud efectiva de esquemas  $\xi = 11$  de longitud
     $l = 2, 3, \dots, 8$  */
int popsize;
    /* Tamaño de la población */
double n_11, n_10, n_01, n_00;
    /* Núm. de los esquemas  $\xi = 11, 10, 01, 00$  en la población */
double P_11, P_10, P_01, P_00;
    /* Probabilidad de seleccionar los esquemas  $\xi = 11, 10, 01, 00$  */
double f_11, f_10, f_01, f_00;

```

```

    /* Aptitud de los esquemas  $\xi = 11, 10, 01, 00$  */
double n_opt[7];
    /*  $n\_opt[l] = esq[l] / (8-l+1)$  para  $l = 2, 3, \dots, 8$  */
double f_opt[7];
    /*  $f\_opt[l] = f\_eff[l] / (8-l+1)$  para  $l = 2, 3, \dots, 8$  */

```

statistics(pop)

```

/* Calcula el núm. de esquemas  $\xi = 11$  de longitud  $l$  y su aptitud efectiva */
struct individual *pop;

```

```

{
    int i, j, k;
        /* contadores */
    double sumfitness;
        /* suma de las aptitudes de los individuos de la población */
    double avg;
        /* aptitud promedio de la población */

    /* Calcula avg */
    sumfitness = 0.0;
    for (j=0; j< popsize; j++)
        sumfitness = sumfitness + pop[j].fitness;
    avg = sumfitness / popsize;

    /* Calcula esq[j] y f_eff[j] para  $j = 0, 1, \dots, 6$  */
    for (i=0; i<7; i++)
    {
        esq[i] = 0;
        f_eff[i] = 0;
        n_opt[j] = 0;
        f_opt[j] = 0;
    }

    for (k=2; k< 9; k++)
    /* Una iteración para cada una de las longitudes  $l = 2, 3, \dots, 8$  */
    {
        for (i=1; i< k; i++)
            /*  $i$  es la posición del primer bit definido en el esquema */
            {
                j = i + (9 - k);

```

```

        /* j es la posición del segundo bit definido en el esquema */
        app_stats(pop, i, j);
        j = 9 - k - 1;
        /* Posición, en el arreglo, de los esquemas de longitud k */

        esq[j] += n_11;
        if (P_11 != 0)
            f_eff[j] += f_11 - ((j+1) / (7*avg)) * ((f_00 * P_00 * f_11)
                - ((f_01 * P_01 * f_10 * P_10) / P_11));
    }
}

/* Calcula n_opt y f_opt */
for (l=0; l<7; l++)
{
    k = l + 2;
    /* Longitud del esquema */

    n_opt[j] = esq[j] / ((double)(8 - k + 1));
    f_opt[j] = f_eff[j] / ((double)(8 - k + 1));
}
}

app_stats(pop, i, j)
/* Calcula n_kl y f_kl con k, l = {0, 1} para esquemas que tienen el primero y
segundo bits definidos en las posiciones i y j respectivamente. */
struct individual * pop;
{
    int k, mask = 1;
    unsigned tmp, tmp1;

    n_11 = 0.0; n_10 = 0.0; n_01 = 0.0; n_00 = 0.0;
    f_11 = 0.0; f_10 = 0.0; f_01 = 0.0; f_00 = 0.0;

    for (k=0; k< popsize; k++)
    {
        tmp = pop[j].chrom;
        tmp1 = tmp >> i - 1;
        if((tmp1 & mask) == 1)

```

```

    {
        tmp1 = tmp >> j - 1;
        if ((tmp1 & mask) == 1) //esquema  $\xi = 11$ 
        {
            n_11 ++ ;
            f_11 = f_11 + pop[j].fitness;
        }
        else /* esquema  $\xi = 10$  */
        {
            n_10 ++ ;
            f_10 = f_10 + pop[j].fitness;
        }
    }
    else
    {
        tmp1 = tmp >> j-1;
        if((tmp1 & mask) == 1) /* esquema  $\xi = 01$  */
        {
            n_01 ++ ;
            f_01 = f_01 + pop[j].fitness;
        }
        else /* esquema  $\xi = 00$  */
        {
            n_00 ++ ;
            f_00 = f_00 + pop[j].fitness;
        }
    }
}

if (n_11 != 0)    f_11 = f_11 / n_11;
if (n_10 != 0)    f_10 = f_10 / n_10;
if (n_01 != 0)    f_01 = f_01 / n_01;
if (n_00 != 0)    f_00 = f_00 / n_00;

P_11 = n_11 / popsize;
P_10 = n_10 / popsize;
P_01 = n_01 / popsize;
P_00 = n_00 / popsize;
}

```

BIBLIOGRAFÍA

- [1] Colin R. Reeves. *Modern Heuristic Techniques for Combinatorial Problems*. John Wiley & Sons, Inc., 1993.
- [2] David E. Goldberg. *Genetic Algorithms in Search, Optimization, and Machine Learning*. Addison Wesley, 1989.
- [3] C.R. Stephens and H. Waelbroeck. *Effective Degrees of Freedom in Genetic Algorithms and the Block Hypothesis*. Proceedings of the Seventh International Conference on Genetic Algorithms, 1997, ed. T. Bäck, 34 (Morgan Kaufman, San Mateo).
- [4] J.M. Yeomans. *Statistical Mechanics of Phase Transitions*. Oxford Science Publications, 1993.
- [5] P. Tarazona. *Error Thresholds for Molecular Quasispecies as Phase Transitions: From Simple Landscapes to Spin-Glass Models*. Physical Review A., 1991, Volume 45, Number 8, P.6038-6050.
- [6] Ira Leuthäusser. *Statistical Mechanics of Eigen's Evolution Model*. Journal of Statistical Physics, Vol. 48, Nos. 1/2, 1987.
- [7] C.R. Stephens and H. Waelbroeck. *Schemata Evolution and Building Blocks*. Evol. Comp., 1998, por publicarse.
- [8] Stuart A. Kauffman. *The Origins of Order*. Oxford Univ. Press, 1993.
- [9] Gabriela Ochoa and Inman Harvey. *Recombination and Error Thresholds in Finite Populations*. FOGA V, 1998.
- [10] Gérard Weisbuch. *Complex Systems Dynamics*. Addison Wesley, 1991.
- [11] John H. Holland. *Adaptation in Natural and Artificial Systems*. MIT Press, 1994.
- [12] C.R. Stephens and H. Waelbroeck. *Analysis of the Effective Degrees of Freedom in Genetic Algorithms*. Phys. Rev., 1998, D57 3251.

-
- [13] Manfred Eigen, John McCaskill and Peter Schuster. *The Molecular Quasi-Species*. Adv. Chem. Phys., 1989 75 149.
- [14] C.R. Stephens, H. Waelbroeck and R. Aguirre. *Schemata as Building Blocks: Does Size Matter?* FOGA V, 1998.
- [15] Paul G. Higgs. *Error Thresholds and Stationary Mutant Distributions in Multi-Locus Diploid Genetics Models*. Genet. Res., Camb. (1994), 63, pp. 63-78.
- [16] Giorgio Parisi. *On The Emergence of Tree-Like Structures in Complex Systems*. Perspectives on Biological Complexity, eds. O.T. Solbrig & G. Nicolis (IUBS, 1991).
- [17] Murray Gell-Mann. *What is Complexity?* Complexity, Vol. 1, Num. 1, 1995, pp. 16-19.