



UNIVERSIDAD NACIONAL  
AUTÓNOMA DE  
MÉXICO

**UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO**

---

---

POSGRADO EN CIENCIA E INGENIERÍA DE LA COMPUTACIÓN

**EMERGENCIA DE GRADOS DE LIBERTAD  
EFECTIVOS EN SISTEMAS ADAPTATIVOS**

**T E S I S**

QUE PARA OBTENER EL GRADO DE:

**DOCTOR EN CIENCIAS  
(COMPUTACIÓN)**

**P R E S E N T A:**

**JORGE CERVANTES OJEDA**

**DIRECTOR DE TESIS: DR. CHRISTOPHER STEPHENS**

**MÉXICO, D.F.**

**2007**



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.

*A mi esposa Carmen*

*A mis hijos Elisa y José Arturo*

*A Dios*

# *Agradecimientos*

*Al Dr. Christopher R. Stephens por brindarme su apoyo permanente, por escucharme, por motivarme y por lo mucho que aprendí de él.*

*A los miembros del comité tutorial, la Dra. Katya Rodríguez Vázquez y el Dr. Pedro Miramontes Valle por el tiempo que dedicaron a la revisión y mejoramiento de este trabajo.*

*Al CONACYT por concederme apoyo económico mediante la beca 165413 y la beca para formación de Doctores en Ciencias.*

*Al COMECYT por su apoyo mediante una beca-tesis.*

*Al posgrado en Ciencia e Ingeniería de la Computación de la UNAM, a la DGEP y al Macroproyecto de la UNAM “Tecnologías para la universidad de la información y la computación” por los apoyos recibidos.*



# Resumen

Esta tesis explica, en forma muy detallada, los Grados de Libertad Efectivos (GLE) emergentes en un modelo de un sistema genético evolutivo, que es un sistema adaptativo o autoadaptable. Los GLE son aquellos que sirven para describir el comportamiento de un sistema de manera simplificada ya sea exacta o aproximadamente. Los GLE emergentes no son visibles sino hasta que las interacciones entre los objetos microscópicos dan lugar a comportamientos macroscópicos cualitativamente diferentes a los que se tienen a nivel microscópico.

El modelo elegido es el de un sistema en el que actúan, sobre una población, operadores genéticos como: la selección (proporcional a la aptitud), la mutación (con tasas fijas) y recombinación (en diversas variantes) sobre diversos paisajes de aptitud que representan situaciones extremas y también sobre algunos paisajes de aptitud más complejos<sup>1</sup>. De este modelo se obtienen conclusiones aplicables tanto a la genética poblacional como a la Computación Evolutiva (concretamente a los Algoritmos Genéticos o AG). Para el análisis se usan elementos teóricos (formulaciones matemáticas con población infinita) y empíricos (simulaciones con población finita) iniciando con la configuración más simple y se va agregando mayor complejidad.

Primero se usan solamente los operadores de selección y mutación en un AG en donde se analizan, con más detalle y/o con un enfoque distinto a la literatura disponible, fenómenos como: tasa crítica de mutación, ruido y universalidad de paisajes de aptitud, degeneración, robustez, *bloat* (aumento en el número de intrones), tasa de transición de fase de primer orden y tasa óptima de mutación para búsqueda genética en diferentes circunstancias. Se concluye que el GLE emergente es la Cuasiespecie.

Más adelante se analizan por separado los GLE del efecto del operador de recombinación solo, para luego hacerlo también en conjunto con la selección y la mutación (el Algoritmo Genético Simple o AGS). Se muestra cómo depende el efecto de la recombinación en función del estado de la población. Se demuestra que el GLE del efecto destructivo o constructivo de la recombinación es el grado de coincidencia entre los puntos de cruzamiento con las fronteras de los Bloques Constructores del paisaje de aptitud. Se demuestra también que, en función del tipo de recombinación usado y del paisaje de aptitud, se induce un rompimiento de simetría entre los fenotipos y se describen los GLE de éste. Se muestra cómo es que, en algunos casos,

---

<sup>1</sup> El concepto de complejidad no está bien definido aún y hay diferentes versiones de lo que ésta es. Unos dicen, por ejemplo, que los sistemas complejos son aquellos de los que se puede decir mucho mientras otros dicen que son aquellos sistemas transdisciplinarios, es decir, los que por su naturaleza requieren de la interdisciplina para poder ser analizados. En este trabajo se asume que la primera de estas dos es válida.

el GLE para salir de óptimos locales es el balance entre selección y mutación y que la recombinación no afecta significativamente.

Se describe luego una nueva variante de algoritmo genético, el *Rank GA*, creado gracias al descubrimiento de los GLE descritos anteriormente, y se demuestra que, en paisajes de aptitud en donde exista una mínima correlación en las señales del paisaje que lleve hacia el óptimo, es superior a un “meta” Algoritmo Genético Simple (con cualquier tasa de mutación fija), con lo que se pone de manifiesto una de las ventajas de conocer los GLE del sistema. Se extrae, además, del *Rank GA* una filosofía de adaptabilidad aplicable también a otras áreas de la computación evolutiva y que también sirve como modelo en sistemas biológicos. El desarrollo de este algoritmo no estaba en los objetivos iniciales del desarrollo de la tesis pero, dada su importancia como modelo teórico y buenos resultados, se incluye como un corolario de las conclusiones a las que se logró llegar.

Para ir más allá de los paisajes de aptitud con situaciones modelo básicas, se analizan los GLE bajo una clase de mapeos genotipo-fenotipo complejos inspirados en el concepto de redes de regulación genética presentando diferentes tipos y niveles de neutralidad y donde cada genotipo (o secuencia lineal) representa una estructura de árbol. Se describe el efecto que, en este caso, tiene la neutralidad en la robustez de cada genotipo y la presencia del fenómeno del *bloat*. Se muestra la emergencia de un “lenguaje algorítmico” como un GLE emergente y que este GLE sirve como modelo para explicar (parcialmente) la evolución del lenguaje hablado. Se compara luego el desempeño del Rank GA contra el “meta” Algoritmo Genético Simple en instancias difíciles de estos mapeos complejos para probar su efectividad en problemas complejos.

Finalmente, para ampliar aún más la complejidad del modelo y hacerlo más afín con un sistema genético natural, se describen los GLE en mapeos fenotipo-aptitud complejos mediante el desarrollo y uso de un simulador de vida artificial. En esta simulación, la aptitud no es asignada a los individuos de manera directa sino que, además de ser consecuencia de los rasgos fenotípicos dados, es consecuencia de la coevolución. Se describe la evolución de los rasgos fenotípicos (presentes en el modelo) en función de los diferentes ambientes posibles y la epistasis entre éstos describiendo los GLE presentes. Se logra observar cómo es que los rasgos fenotípicos preferidos por la evolución no son siempre los que le dan más ventaja a los individuos sino a la especie en su conjunto.

En general esta tesis debe servir para quienes estudian teoría de la CE y quieren entender mejor los porqués en los algoritmos genéticos, los que estudian genética poblacional en sistemas biológicos o también para diseñadores de AGs que quieren saber cómo obtener un algoritmo más eficiente para alguna clase de problemas y para usuarios de AGs que deben elegir entre varios algoritmos el más adecuado a sus fines.

# Contenido

<i>Abreviaciones</i>	13
<i>Simbología</i>	15
<i>Introducción</i>	17
<b>Antecedentes</b>	18
<b>Objetivos</b>	20
<b>Herramientas</b>	21
<b>Organización de la tesis</b>	22
<b>Capítulo 1</b> <i>Evolución con selección y mutación – Cuasiespecies</i>	25
<b>1.1 La ecuación de evolución como herramienta de análisis</b>	26
<b>1.2 Casos extremos</b>	28
1.2.1 Operadores con efecto nulo	28
1.2.2 Cuando la probabilidad de mutación es 1/2	28
1.2.3 Población uniformemente distribuida y en equilibrio	28
1.2.4 Paisaje de aptitud plano y población en equilibrio	29
1.2.5 Solución de la ecuación de evolución para un gen ( $N = 1$ )	30
1.2.6 Solución general de la ecuación de evolución	35
<b>1.3 La Cuasiespecie en el paisaje de aguja en un pajar</b>	36
1.3.1 La “forma” de la Cuasiespecie	37
1.3.1.1 La Cuasiespecie cuando $p = 0$	39
1.3.1.2 La Cuasiespecie cuando $p = 1$ :	40
1.3.2 La tasa crítica de mutación de una Cuasiespecie	42
1.3.2.1 Aproximaciones teóricas para la tasa crítica de mutación $p^*$	42
1.3.2.2 Tasa de mutación crítica y Aptitud Efectiva de la secuencia maestra	46
1.3.2.3 Tasa crítica de mutación $p^*$ como función de la longitud de los genomas $N$	49
1.3.2.4 Aguja en un pajar con ruido. Universalidad del paisaje de aguja en un pajar	51
<b>1.4 La Cuasiespecie y agujas degeneradas</b>	53
1.4.1 Degeneración con secuencias maestras contiguas	54
1.4.1.1 Tasa de mutación crítica en función del grado de degeneración de la aguja	56
1.4.1.2 La formación de una Cuasiespecie a través del tiempo	58
1.4.1.3 Poblaciones con genotipos de longitudes diferentes	64

1.4.1.4	Mutación capaz de modificar la longitud de los genomas	67
1.4.2	Degeneración con secuencias maestras separadas	69
<b>1.5</b>	<b>Competencia entre Cuasiespecies</b>	<b>71</b>
1.5.1	Transiciones de fase de segundo orden	71
1.5.2	Transiciones de fase de primer orden	72
<b>1.6</b>	<b>Cuencas de atracción</b>	<b>80</b>
<b>1.7</b>	<b>Cuasiespecies con población finita</b>	<b>82</b>
1.7.1	Búsqueda de una aguja en un pajar	82
1.7.1.1	Tasa de mutación óptima para encontrar el óptimo	82
1.7.1.2	Tasa de mutación y tamaño de la población óptimas para maximizar la probabilidad de supervivencia de la secuencia maestra	83
1.7.2	Búsqueda en paisajes con dos señales	85
1.7.2.1	Subóptimo anticorrelacionado	85
1.7.2.2	Cooperación o decepción	86
1.7.2.3	Búsqueda de una señal desde la otra	88
<b>1.8</b>	<b>Conclusiones</b>	<b>89</b>
	<b>Notas</b>	<b>93</b>
<i>Capítulo 2</i>	<i>Evolución con recombinación ? Bloques Constructores</i>	<i>95</i>
<b>2.1</b>	<b>Teorema de esquemas y la hipótesis de bloques constructores</b>	<b>96</b>
<b>2.2</b>	<b>La ecuación de evolución exacta con recombinación para secuencias genéticas</b>	<b>97</b>
<b>2.3</b>	<b>La ecuación de evolución exacta con recombinación para esquemas</b>	<b>98</b>
<b>2.4</b>	<b>Métricas</b>	<b>100</b>
2.4.1	Distancia al centro del simplex	100
2.4.2	Distancia a la variedad de Geiringer	101
2.4.3	Distancia al óptimo	101
2.4.4	Frecuencia relativa por efecto de la recombinación	102
<b>2.5</b>	<b>Paisajes de aptitud modulares</b>	<b>102</b>
<b>2.6</b>	<b>Efecto de la recombinación sola</b>	<b>103</b>
2.6.1	Teorema de Geiringer	103
2.6.2	Resultados	104
<b>2.7</b>	<b>Efecto de la recombinación con mutación</b>	<b>105</b>
<b>2.8</b>	<b>Efecto de la recombinación con selección</b>	<b>108</b>
2.8.1	Paisaje de 1 bloque de Aguja en un pajar	108
2.8.2	Paisaje de N bloques de Aguja en un pajar	110
2.8.3	Dos bloques concatenados de Aguja en un pajar	113
<b>2.9</b>	<b>Población finita con recombinación, selección y mutación</b>	<b>115</b>
2.9.1	Aguja en un pajar	115
2.9.2	Concatenación de paisajes de Aguja en un pajar	118

2.9.2.1	Concatenación de 3 paisajes de Aguja en un pajar de 4 bits cada uno	118
2.9.2.2	Paisaje con N bloques concatenados de aguja en un pajar	120
2.9.3	Trampa o paisaje engañoso	121
2.9.4	Trampas concatenadas	122
2.9.5	Dos señales anticorrelacionadas	123
2.9.6	Dos señales correlacionadas	124
<b>2.10</b>	<b>Conclusiones</b>	<b>127</b>
	<b>Notas</b>	<b>129</b>
<b>Capítulo 3</b>	<b><i>Operadores genéticos dependientes del Ranqueo</i></b>	<b>131</b>
<b>3.1</b>	<b>El Rank GA</b>	<b>132</b>
3.1.1	Mutación proporcional al Ranqueo	133
3.1.2	Recombinación con apareamiento de Ranqueo cercano	134
3.1.3	Evaluaciones intermedias	135
<b>3.2</b>	<b>Resultados</b>	<b>135</b>
3.2.1	Concatenación de N bloques de aguja en un pajar de 1 bit	136
3.2.2	Trampa engañosa	137
3.2.3	Dos señales anticorrelacionadas	138
3.2.4	Dos señales correlacionadas	139
3.2.5	Bloques concatenados	141
3.2.5.1	Bloques concatenados y puntos de cruzamiento limitados	143
<b>3.3</b>	<b>Conclusiones</b>	<b>145</b>
	<b>Notas</b>	<b>147</b>
<b>Capítulo 4</b>	<b><i>Mapeos Genotipo-Fenotipo complejos con neutralidad</i></b>	<b>149</b>
<b>4.1</b>	<b>Neutralidad e intrones</b>	<b>150</b>
<b>4.2</b>	<b>Un mapeo genotipo-fenotipo complejo con neutralidad inspirado en las Redes de Regulación Genéticas</b>	<b>151</b>
<b>4.3</b>	<b>Efecto de la neutralidad</b>	<b>153</b>
<b>4.4</b>	<b>Emergencia de un lenguaje algorítmico</b>	<b>155</b>
<b>4.5</b>	<b>Búsqueda con operadores dependientes del Ranqueo</b>	<b>157</b>
4.5.1	Casos con alta neutralidad	158
4.5.2	Casos con menos neutralidad	159
<b>4.6</b>	<b>Conclusiones</b>	<b>160</b>
	<b>Notas</b>	<b>161</b>
<b>Capítulo 5</b>	<b><i>Mapeo Fenotipo-Aptitud complejo</i></b>	<b>163</b>
<b>5.1</b>	<b>El Sistema</b>	<b>163</b>
<b>5.2</b>	<b>El modelo</b>	<b>164</b>
<b>5.3</b>	<b>El Simulador</b>	<b>165</b>
5.3.1	Terreno	165

---

5.3.2	Alimento	166
5.3.3	Especies	168
<b>5.4</b>	<b>Resultados</b>	<b>168</b>
5.4.1	Individuos solos	169
5.4.2	Especies puras. 30 individuos iguales	172
5.4.3	Misma estrategia con diferentes capacidades	175
5.4.3.1	Diferente capacidad sensorial, misma capacidad metabólica	175
5.4.3.2	Diferente capacidad metabólica, misma capacidad sensorial	177
5.4.3.3	Diferente capacidad sensorial y metabólica en cada individuo	180
5.4.4	Estrategia diferente, misma capacidad sensorial y metabólica	182
5.4.5	Estrategia diferente, diferente capacidad sensorial y metabólica	184
<b>5.5</b>	<b>Conclusiones</b>	<b>186</b>
	<b>Notas</b>	<b>189</b>
<i>Capítulo 6</i>	<i>Conclusiones Finales</i>	<i>191</i>
<i>Apéndice 1:</i>	<i>Simetrías en la Integración de la Ecuación de Evolución</i>	<i>193</i>
<i>Apéndice 2:</i>	<i>Curvas Características de Búsqueda</i>	<i>195</i>
<i>Apéndice 3:</i>	<i>Opciones del simulador</i>	<i>197</i>
<i>Apéndice 4:</i>	<i>Población para el ejemplo del capítulo 4.</i>	<i>203</i>
	<i>Referencias</i>	<i>207</i>
	<i>Notas</i>	<i>213</i>

# Abreviaciones

<b>Español</b>	<b>Inglés</b>	<b>Descripción</b>
AG	GA	Algoritmos Genéticos
CE	EC	Computación Evolutiva
GLE	EDoF	Grados de Libertad Efectivos
HBC	BBH	Hipótesis de Bloques Constructores
OMOE	EMOO	Optimización Multiobjetivo Evolutiva
PG	GP	Programación Genética
SAC	CAS	Sistemas Adaptativos Complejos
TEH	HST	Teorema de Esquemas de Holland
TIC	ICT	Tecnologías de la Información y de la Comunicación
sm	ms	secuencia maestra
nsm	nms	no secuencia maestra o secuencia no maestra

# Simbología

$\beta$	Bloque del paisaje
${}^N C_q$	número de combinaciones posibles de $N$ objetos tomados de $q$ en $q$ .
$d_{ij}$	distancia de Hamming entre $i$ y $j$
$f$	aptitud
$\bar{f}$	aptitud promedio
$f_{eff}$	aptitud efectiva
$G$	ganancia de aptitud
$i, j, k$	secuencias genéticas
$i \rightarrow j$	evento de que $i$ mute en $j$
$\bar{i}$	secuencia genética opuesta a la secuencia $i$
$I, J, K$	esquemas
$I_m$	esquema formado con los bits de $I$ indicados como primer padre en $m$
$L$	número mínimo de bits en 1 para que una secuencia sea maestra
$m$	máscara de recombinación
$\bar{m}$	máscara opuesta a la máscara $m$
$M$	número de bloques en el paisaje de aptitud
$n$	longitud del genoma cuando hay diferentes longitudes simultáneas
$N$	longitud del genoma o longitud máxima de los genotipos



conjunto de los números naturales incluyendo al cero  $\{0, 1, 2, \dots\}$

$P$  concentración en la población como fracción,  $0 = P = 1$ .

$(j \rightarrow i)$  probabilidad de que  $j$  mute en  $i$

$p$  tasa de mutación

$p^*$  tasa de mutación crítica

$p_f$  tasa de mutación de transición de fase de primer orden

$p_c$  probabilidad de cruzamiento

$p_m$  probabilidad de usar la máscara  $m$

conjunto de los números reales.

$t$  número de generación

$x_i$  número de individuos con la secuencia genética  $i$

$\propto$  es proporcional a

$\approx$  es aproximadamente igual a

# Introducción

Los sistemas adaptativos se caracterizan por su capacidad para autoadaptarse a situaciones no previstas por lo que, hoy en día, son cada vez más usados en ambientes con complejidad creciente y, por lo tanto, presentan un comportamiento “no lineal” que es difícil de explicar. Un ejemplo típico de un sistema adaptativo es la evolución de las especies en la naturaleza la cual ha inspirado a los investigadores a crear la llamada Computación Evolutiva (CE) y, dentro de ésta, los Algoritmos Genéticos (AG), los cuales han tenido mucho éxito y que buscan imitar la manera en que la naturaleza ha logrado encontrar “soluciones” genéticas a las condiciones cambiantes que la vida en el planeta ha tenido que enfrentar. Pero ¿cómo funcionan los AG? ¿bajo qué circunstancias son exitosos? ¿puede explicarse la genética poblacional en la naturaleza modelándola con un AG? son preguntas que en cierta medida siguen abiertas.

Todo sistema tiene un número de grados de libertad. Estos grados de libertad son las posibles dimensiones que en el sistema pueden variar como, por ejemplo, la posición, velocidad, aceleración, velocidad angular, temperatura, etc. y que son necesarias para determinar el estado del mismo. A diferencia de sistemas más simples donde los grados de libertad “microscópicos” se suman para formar grados de libertad “macroscópicos”, los cuales exhiben también un comportamiento relativamente simple, los sistemas adaptativos complejos (CAS por sus siglas en inglés) tienen un comportamiento radicalmente diferente a diferentes escalas, exhibiendo una rica jerarquía de grados de libertad “emergentes” en cada nivel, que son cualitativamente diferentes. Se dice que estos grados de libertad son emergentes porque no son visibles sino hasta que surgen a partir de las interacciones entre los objetos a nivel microscópico.

Los grados de libertad efectivos (GLE), son aquellos que resultan más prácticos para describir, de manera exacta o aproximada, el comportamiento de un sistema y pueden ser microscópicos y/o macroscópicos. Por ejemplo, cuando se pretende predecir el clima, aunque en el sistema hay grados de libertad como la posición, masa, velocidad, composición química, temperatura, etc. de cada partícula en la atmósfera, resulta imposible describir el comportamiento global a ese nivel. Para esto es preferible usar los GLE como son: la época del año, hora del día, nubosidad, presión atmosférica, etc. para así llegar a una descripción más práctica e intuitiva del sistema.

¿Cómo emergen grados de libertad efectivos complejos a partir de un sistema que a nivel microscópico es simple? es, quizás, la pregunta más importante que hay que enfrentar, tanto en el estudio de sistemas complejos adaptativos como en el de los no adaptativos.

Son ejemplos de sistemas adaptativos complejos: los mercados financieros; en biología, las redes de regulación genética, importantes por ejemplo para la comprensión de enfermedades

asociadas con expresiones anormales de un gen; los sistemas de seguridad, estabilidad y enrutamiento óptimo de tráfico en redes de telecomunicaciones. Dada la creciente importancia y aplicabilidad en ambientes cada vez más complejos de los sistemas adaptativos, se prevé que la ciencia del siglo XXI será en buena medida la ciencia de los Sistemas Adaptativos Complejos (CAS).

El estudio de los CAS es necesariamente transdisciplinario haciéndose necesario el punto de vista de diversas especialidades, entre ellas la de la computación evolutiva (CE). Este estudio, desde el punto de vista de la CE, puede hacerse mediante modelos matemáticos pero también requiere el uso intensivo de simulaciones por computadora ya que estos modelos matemáticos son casi siempre intratables de manera analítica y también porque, para muchos de estos sistemas, no es posible explorar fácilmente casos de prueba *in situ*.

En esta tesis se describe la emergencia de GLEs en modelos matemáticos y simulaciones por computadora de sistemas genéticos evolutivos (que son una forma de sistemas adaptativos complejos) usando tanto elementos teóricos como empíricos. Con el presente trabajo se logra un mejor entendimiento de la respuesta de un sistema genético evolutivo bajo diversas circunstancias mediante la identificación de los GLE emergentes. Se inicia desde la configuración más simple y se van agregando elementos que lo van haciendo más complejo. De esta manera se identifican los GLE presentes debido a la acción de cada uno de estos elementos agregados y también los que se deben a las interacciones entre ellos.

## Antecedentes

Una de las áreas más interesantes de los sistemas adaptativos complejos es la dinámica genética, que es de mucha importancia no solamente en la biología poblacional [ 11 ] sino también en la computación evolutiva (CE) [ 6 ][ 30 ], siendo la última inspirada por la primera. En ambas disciplinas, una población de “objetos” evoluciona bajo el efecto de un conjunto de operadores genéticos que actúan estocásticamente, siendo los más comunes la selección, la mutación y la recombinación. La selección da un sesgo hacia una mayor presencia de los objetos con mayor “aptitud”, lo cual se mide a través de una función de aptitud (*fitness*) que depende de la configuración genética de cada objeto. La mutación actúa aleatoriamente para modificar materia genética de un objeto y llevarlo así de un estado (genético) a otro. La recombinación toma materia genética de dos objetos “padres” y la mezcla para formar objetos “hijos” nuevos. Para explicar esta dinámica, uno de los conceptos fundamentales que han surgido en la teoría de sistemas genéticos es el de un paisaje de aptitud o *fitness landscape* [ 73 ][ 53 ] donde la aptitud está definida como una “función de altura” por encima del espacio de configuraciones (genéticas). El concepto de evolución, en este contexto, se entiende como un proceso donde la población, mediante la acción de los operadores genéticos, trepa por las montañas de esta topografía rugosa en busca de los mejores valores de aptitud (los más altos).

Los sistemas genéticos en los que actúan solamente la selección y la mutación han sido estudiados en [ 19 ] en el marco de la Biología Molecular. En dicho trabajo se analizan, entre

otras cosas, los puntos de equilibrio del sistema en el límite de una población infinita en función del grado de exactitud en la replicación de las moléculas (análogo de la tasa de mutación). Se describe la existencia del fenómeno de *transiciones de fase de primer orden* de los puntos de equilibrio en función de la tasa de mutación (o exactitud de replicación de las moléculas), sin embargo, no se analizan las dependencias de estas transiciones de fase ni la fase transitoria del sistema ni sus implicaciones en poblaciones finitas.

En sistemas recombinativos, es decir, en donde la recombinación es el principal operador de búsqueda genética (más importante que la mutación), la dinámica del sistema es mucho más compleja y por esto es que se les ha dedicado más esfuerzo para entenderlos en la teoría de la CE. Los primeros fundamentos teóricos de la CE, en el contexto de los AGs, fueron el teorema de esquemas [ 30 ] y la hipótesis de bloques constructores [ 25 ]. El teorema de esquemas establece que la fracción de la población que ocupará un esquema en la siguiente generación tiene una cota inferior y se da una expresión matemática para esa cota. La hipótesis de bloques constructores dice que un AG encuentra una buena solución al combinar esquemas (subconjuntos de bits de las cadenas) aptos y cortos. Estos elementos proporcionaron un marco intuitivo dentro del cual resultó posible empezar a entender cómo funcionan los AGs. Sin embargo, esta teoría ha sido atacada en la literatura [ 15 ][ 2 ][ 21 ][ 22 ]. El defecto principal del teorema de esquemas es que no toma en cuenta explícitamente el efecto de reconstrucción de esquemas (o cadenas) a través del cruzamiento. En el caso de la hipótesis de bloques constructores, aunque es intuitivo, no cuenta con mucha evidencia experimental, aún en problemas creados para probar su validez [ 23 ].

Este tipo de tratados ha sido inspirado por la idea de encontrar una teoría aproximada que ofrezca intuición y que pueda hacer contacto con problemas “reales”. En esta dirección está el trabajo de Shapiro y colaboradores [ 52 ] quienes escribieron una teoría efectiva basada en métodos de la mecánica estadística para la evolución de la distribución de probabilidades para un AG. Cabe mencionar que ni en la biología poblacional ni en la CE existen métodos generales de aproximación con los cuales las ecuaciones puedan ser resueltas.

Otra metodología teórica que ha sido usada para analizar la dinámica genética se basa en el formalismo de las cadenas de Markov [ 68 ][ 69 ][ 70 ]. Las aplicaciones de éstas han sido principalmente a AGs canónicos con cadenas de longitud fija pero también se han extendido al caso de cadenas de longitud variable [ 51 ]. En estos casos se modela la evolución de una población como una cadena de Markov y se usan sus propiedades para llegar a resultados formales sobre su evolución, en particular, sobre las limitantes de las distribuciones de la población en el límite de tiempo largo. Este formalismo tiene la ventaja de que es una teoría exacta y rigurosa; sin embargo es microscópico y muy difícilmente ofrece un entendimiento cualitativo sobre el comportamiento del sistema.

Recientemente ha surgido una teoría que combina los puntos buenos de los métodos mencionados [ 56 ][ 62 ][ 63 ][ 64 ]. Esta teoría está basada en un granulado grueso (*coarse graining*) de los grados de libertad microscópicos, introduciendo grados de libertad “efectivos”. Como la teoría de cadenas de Markov, es una teoría con un fundamento microscópico y riguroso pero, sin embargo, no ilustra mucho la interpretación intuitiva de la dinámica genética, y en particular el efecto de la recombinación. Esta teoría da lugar a un teorema de esquemas que explícitamente incluye el efecto de reconstrucción de esquemas a

través del cruzamiento planteando una ecuación de evolución exacta en lugar de una desigualdad como en el caso de la teoría de Holland. La forma de la ecuación también sugiere inmediatamente una interpretación en términos de construcción de soluciones a través de bloques constructores. Sin embargo, a diferencia de la hipótesis estándar de bloques constructores, esta teoría muestra que, bajo ciertas circunstancias, son los esquemas largos y no los cortos los que dominan [ 65 ].

Con esta teoría de granulado grueso en la CE ha sido posible llegar a resultados nuevos en la biología poblacional, como una extensión del teorema de Geiringer [ 24 ] a representaciones lineales de longitud variable y a representaciones de tipo árbol usando tanto recombinación homóloga como de tipo “subárbol”.

Por otro lado, la naturaleza de los objetos que evolucionan depende mucho del problema específico de interés. Frecuentemente en la biología poblacional un objeto es un par de cromosomas que se modelan a través de cadenas de bits, cada bit representando un *locus* o posición dentro del material genético cuyo estado puede tomar uno de dos valores posibles, 0 ó 1. En la Programación Genética (PG) un objeto es una estructura en forma de árbol que representa a un programa. En principio, representaciones aún más generales pueden ser consideradas como redes neuronales. El cómo actúan los operadores genéticos depende de qué representación se use para los objetos. El que dichas representaciones en las ramas principales de la Computación Evolutiva (CE) - Algoritmos Genéticos (AGs) [ 30 ][ 25 ], programación Genética (PG) [ 37 ][ 38 ] y Estrategias Evolutivas (EE) [ 8 ][ 6 ] - sean tan diferentes, ha propiciado que se desarrollen diferentes teorías para cada una. Este hecho oculta la esencia única de la dinámica genética que va más allá de la representación de los objetos cuya dinámica se pretende describir. La teoría de granulado grueso ha sido extendida al caso de la PG recientemente [ 47 ][ 48 ][ 50 ], mostrando así que sus ventajas, en relación con otras formulaciones, se extienden a otras áreas de la CE, específicamente, a objetos tales como árboles.

## Objetivos

- ? Desarrollar un mejor marco formal para el entendimiento de problemas dinámicos usando los conceptos de aptitud efectiva (*effective fitness*) y rompimiento inducido de simetría (*induced symmetry breaking*).
- ? Estudiar la teoría de la Dinámica Genética en términos de una formulación de tipo *coarse grained* (de granulado grueso), basada en ecuaciones dinámicas exactas.
- ? Validación de los desarrollos planeados en los campos de aplicación particulares relacionados: la CE y la biología poblacional.
- ? Usar el concepto de aptitud efectiva para entender mejor fenómenos importantes como *bloat* (expansión), redes neutrales, robustez de las soluciones ante los operadores genéticos.
- ? Examinar la emergencia de lenguajes algorítmicos simples en mapeos genotipo-fenotipo complejos que modelan redes de regulación genética.

- ? Estudiar mapeos fenotipo-aptitud complejos similares a los que están presentes en la naturaleza.

## Herramientas

La herramienta teórica principal que se usa es la desarrollada en [ 62 ][ 63 ][ 64 ] y, en el contexto de la PG, en [ 47 ][ 48 ] la cual ha llevado a la llamada “Stephens and Waelbroeck's GA Schema Theory” [ 38 ]. En esta teoría se pasa de una representación en términos de los objetos fundamentales (cadenas, árboles, etc.) a una representación en términos de objetos de granulado grueso (*coarse grained*) – generalmente esquemas – llegando a ecuaciones exactas que describen la evolución de cualquier esquema bajo los tres operadores principales de selección, mutación y recombinación. La dinámica de un esquema dado es descrita en términos de la evolución de esquemas de orden más bajo (de granulado más grueso). Así se forma una jerarquía donde puede verse cómo se construye un esquema dado en términos de sus bloques constructores. Se puede formular la transformación de la base microscópica a la de grano grueso mediante una transformación de coordenadas en el espacio de configuraciones del modelo. Esta nueva base – la base de bloques constructores – tiene muchas propiedades interesantes.

Esta jerarquía termina en los esquemas de orden uno que ya no pueden dividirse a través de recombinaciones. Estos forman los bloques constructores de más bajo nivel. Así, las ecuaciones capturan de forma natural la idea de que un objeto apto se encuentra mediante la yuxtaposición de objetos de más bajo nivel o de granulado más grueso (esquemas de menor orden) aptos.

Otra herramienta teórica que se usará es el concepto de “aptitud efectiva” y el de “rompimiento inducido de simetría”. La aptitud efectiva se define como aquel valor de aptitud selectiva en una dinámica genética más general en la que se descartan otros operadores genéticos que llevaría al mismo número de descendientes que en la dinámica con varios operadores. Es decir, se considera una dinámica con varios operadores genéticos (selección, mutación y recombinación) y se calcula, dado un cierto número de objetos de un cierto tipo (por ejemplo un esquema) en la generación  $t$ , ¿cuántos habrá de éstos en la generación  $t+1$ ? Dado este número, se asume que solamente hay selección y se calcula el valor de la aptitud en la generación  $t$  que llevaría al mismo número de estos objetos en la siguiente generación. El paisaje que corresponde a la aptitud efectiva es dinámico dado que la aptitud efectiva depende del estado actual de la población. En muchas situaciones, cuando otros operadores además de la selección juegan un papel importante, la evolución no se puede interpretar como un proceso de *hill climbing* en un paisaje de aptitud. En estos casos, la aptitud efectiva restaura la noción de *hill climbing* aún en la presencia de operadores distintos a la selección. Un ejemplo concreto, donde no se puede entender la evolución en términos de *hill climbing*, es cuando dos genotipos a los que les corresponde el mismo fenotipo, y por lo tanto tienen la misma aptitud, son muy diferentes en términos de sus propiedades bajo la acción de la mutación en caso de que uno sea más robusto que el otro. Se encuentra que la aptitud efectiva del más robusto es mayor que la del otro y por lo tanto habrá a más descendientes de ese genotipo.



El concepto de rompimiento inducido de simetría [ 3 ] se refiere a la distinción que se ejerce mediante los operadores genéticos entre objetos sinónimos, es decir, con diferentes genotipos pero el mismo fenotipo y por lo tanto misma aptitud. Esta distinción se debe a que dichos objetos tienen diferentes resistencias a los efectos de la mutación, afectando o beneficiando más a uno que a otro. Debido a este rompimiento de simetría inducido por la acción de los operadores genéticos se producen cromosomas estructuralmente más estables lo cual permite la exploración de posibles nuevas mejoras genéticas sin caer en el problema de la fragilidad genética (*brittleness problem*).

La integración numérica y la simulación por computadora son herramientas que también se usan masivamente para obtener resultados para validar el análisis teórico y también para analizar casos complejos en donde las ecuaciones resultan intratables analíticamente. Para mostrar muchos de estos resultados se usan las Curvas Características de Búsqueda, definidas en el Apéndice 2 de esta tesis, las cuales proporcionan un marco adecuado para analizar y comparar el desempeño de algún algoritmo de búsqueda dado sobre paisajes de aptitud específicos.

## Organización de la tesis

En el Capítulo 1 se estudia detalladamente la evolución en un sistema genético (un algoritmo genético simple) en el que operan solamente la selección y la mutación, y se hace una descripción de los grados de libertad efectivos presentes. Para esto se describen situaciones básicas extremas y se analizan las interacciones presentes entre los operadores genéticos para luego describirlas en términos de los grados de libertad efectivos. Se analizan, en forma teórica y/o empírica, fenómenos como la tasa crítica de mutación, degeneración, universalidad en paisajes de aptitud similares, tasa de transición de fase, cuencas de atracción y el llamado *bloat* (o hinchazón). En el contexto de la búsqueda genética se describe cual es la tasa de mutación óptima en cada caso y de qué depende.

En el Capítulo 2 se describen los grados de libertad efectivos cuando solamente actúa el operador de recombinación (los bloques constructores). Se analiza luego el sistema con selección, mutación y recombinación actuando en conjunto. Se aplica la teoría de Stephens y Waelbroeck para hacer la primera parte del análisis y luego se usan simulaciones para ver de qué manera afecta la presencia de la recombinación junto con la selección y mutación.

El Capítulo 3 se dedica a describir y poner a prueba un algoritmo genético original (del autor de esta tesis), el *Rank GA*, en el que se han modificado los operadores de mutación y recombinación para dirigir la búsqueda genética de una manera más robusta y justificando dichas modificaciones en términos de los GLE estudiados en los capítulos anteriores. Se hace una comparación del desempeño de este algoritmo contra un “meta” algoritmo genético simple mediante paisajes de aptitud modelo que representan situaciones extremas y que se presentan en aplicaciones prácticas mostrándose las buenas propiedades del algoritmo propuesto.

En el Capítulo 4 se analiza un mapeo genotipo-fenotipo muy rico y complejo, basado en el

concepto de Redes Regulatorias Genéticas, que presenta características que lo hacen interesante como: decepción (señales engañosas), epistasis, modularidad, rugosidad (múltiples picos o ruido) y neutralidad a diferentes niveles. Con este modelo se estudian algunos aspectos del efecto de la neutralidad en la evolución así como una forma alternativa de estudiar sistemas complejos mediante el concepto de “lenguaje algorítmico” emergente en los genotipos de los individuos y sus reglas gramaticales. También se usa este modelo para poner a prueba el algoritmo presentado en el Capítulo 3 ya que se pueden generar instancias de problemas realmente muy difíciles y evaluar su desempeño contra el “meta” algoritmo genético simple.

El Capítulo 5 cierra este estudio con un análisis para describir los GLE en mapeos, ahora, fenotipo-aptitud complejos que agregan aun mayor complejidad al sistema genético ya que la aptitud es asignada a los individuos de manera indirecta y depende de lo que vayan haciendo los demás individuos (coevolución), de manera similar a lo que se ve en la naturaleza. Aquí el estudio se hace mediante el desarrollo de un simulador de vida artificial en el que los individuos representan “animales” coexistiendo y compitiendo entre ellos por una cantidad de alimento finita que está distribuida sobre un terreno también finito.

El Capítulo 6 contiene las conclusiones finales de la tesis.



# Capítulo 1 Evolución con selección y mutación – Cuasiespecies

Los algoritmos genéticos son de mucha importancia e interés en la Computación Evolutiva desde hace ya muchos años dada su adaptabilidad y eficacia para resolver problemas de búsqueda en espacios donde es difícil encontrar una heurística para encontrar soluciones óptimas. Sin embargo resulta siempre difícil, por falta de un entendimiento claro de los grados de libertad efectivos presentes, encontrar los parámetros del algoritmo que resuelven un problema específico de la manera más eficiente y eficaz como la tasa de mutación o el tamaño de la población o la representación más adecuada, etc. [ 26 ][ 4 ]. Por ejemplo, para la tasa de mutación ideal se ha escrito 1) que la tasa  $1/N$  es mejor [ 41 ] ( $N$  es la longitud del genoma), 2) que la tasa ideal cambia y por lo tanto se debe usar una tasa dinámica [ 5 ][ 29 ][ 59 ][ 55 ][ 20 ], y 3) que la tasa ideal depende de la tasa crítica de mutación o umbral de error de replicación [ 44 ][ 45 ][ 43]. La idea es siempre generar heurísticas útiles para determinar el valor de la tasa de mutación que mientras más universales sean es mejor. Sin embargo, la heurística  $1/N$  es demasiado universal ya que no depende del paisaje de aptitud. El umbral de error sí depende del paisaje de aptitud y se le ha asociado con un balance óptimo entre exploración y explotación, sin embargo en la práctica no es posible calcular este umbral ya que no se conocen los detalles del paisaje.

El análisis hecho en este capítulo servirá para distinguir cuáles grados de libertad efectivos en los algoritmos genéticos son consecuencia de la selección y de la mutación por separado y cuales se deben a la interacción entre éstos. El principal antecedente del estudio de este sistema es el de Eigen-McCaskill-Schuster [ 19 ] el cual analiza un sistema con una población infinita de individuos (moléculas) que evolucionan mediante *selección* y *mutación* únicamente. Es decir, que cada individuo se reproduce de manera asexual a una cierta tasa llamada *aptitud reproductiva* (fenotipo), la cual depende directamente de su *secuencia genética* (genotipo) y esta reproducción se hace con una cierta *tasa de error* (probabilidad de mutación) fija que provoca alteraciones genotípicas o *mutaciones* en algunos de sus descendientes. En ese trabajo se describe la emergencia de una Cuasiespecie formada por individuos que no comparten exactamente la misma información genética sino que, en el espacio de los genomas, se agrupan alrededor de un genotipo cuya aptitud es superior a la de los demás. También se muestra la existencia de una transición de fase de primer orden respecto a la tasa de mutación en la distribución final de la población sobre este modelo.

En este capítulo se estudia el mismo sistema (que evoluciona mediante selección y mutación) de una manera aun más detallada haciendo énfasis en la dinámica a través del tiempo en diferentes circunstancias y en la manera en que los detalles del paisaje de aptitud y la tasa de mutación afectan a ésta. De este estudio se pretende crear una intuición, a partir de la definición de los grados de libertad efectivos presentes, que permita prever cualitativamente el comportamiento del sistema bajo cualquier situación.

## 1.1 La ecuación de evolución como herramienta de análisis

Los sistemas evolutivos, aún en la forma más simple que es la que se estudia en este capítulo, tienen un comportamiento que resulta difícil de entender cualitativamente y, por lo tanto, se les debe estudiar de diferentes ángulos y con diferentes metodologías para esclarecer lo más que se pueda la forma en que realmente funcionan. Una de estas formas es mediante un modelo matemático teórico determinístico que permita hacer cálculos de la respuesta esperada del sistema bajo diferentes circunstancias. Para lograr este modelo, debido a que el proceso evolutivo es estocástico dando lugar al fenómeno de la *deriva genética*, es necesario considerar que se tiene una población infinita. La idea es distinguir los sesgos intrínsecos que generan en la población los operadores genéticos cada uno por separado y cuando actúan en conjunto sin deriva genética ya que ésta puede confundir y llevar a conclusiones erróneas.

En el límite de una población infinita, la ecuación que modela un sistema genético que evoluciona mediante selección y mutación, en el que todos los genotipos tienen la misma longitud y en donde cada gen puede tener solamente una de dos configuraciones posibles, es (ver [ 19 ] ó [ 58 ])

$$P_i(t+1) = \sum_{j=1}^{2^N} p^{?_{ij}} (1-p)^{N-?_{ij}} \frac{f_j}{\bar{f}(t)} P_j(t) \quad (1.1)$$

donde:

$t$  es el número de generación (tiempo discreto).  $t = 0$ .

$P_i(t)$  es la **fracción de la población** cuyo genotipo es  $i$  en la generación  $t$ .  $\sum_{i=1}^{2^N} P_i(t) = 1, \forall t$ .

$p$  es la probabilidad de mutación de cada gen.  $0 = p = 1$ .

$?_{ij}$  es la **distancia de Hamming** entre los genotipos  $i$  y  $j$ , es decir, el número de genes en que difieren los genotipos  $i$  y  $j$ .  $0 = ?_{ij} = N$ .

$N$  es el **número de genes** en cada genotipo.

$f_j$  es el valor de **aptitud selectiva** del genotipo  $j$  (*selective fitness*).  $f_j = 0$ . Estos valores conforman el **paisaje de aptitud** sobre el espacio de las posibles configuraciones genéticas de los individuos.

$\bar{f}(t) = \sum_{i=1}^{2^N} f_i P_i(t)$  es la **aptitud promedio** de toda la población en la generación  $t$ .

Para entender mejor esta ecuación se define lo siguiente:

$$(j \rightarrow i) p_{ij}^{2^t} (1-p)^{N-2^t}$$

es la probabilidad de que el genotipo  $j$  mute y se convierta en el genotipo  $i$  lo que implica que  $2^t$  genes muten y que  $N-2^t$  genes no muten;

$$\tilde{f}_j(t) = \frac{f_j}{f(t)}$$

es la **aptitud relativa** (a la población) del genotipo  $j$  en la generación  $t$ , o **ventaja reproductiva** del genotipo  $j$  sobre los demás genotipos presentes en la generación  $t$ ;

$$P'_j(t) = \tilde{f}_j(t) P_j(t)$$

es  $P_j$  después de aplicar solamente selección ó la fracción de la población que ocuparán los descendientes de los individuos con genotipo  $j$  presentes ó la probabilidad de seleccionar el genotipo  $j$  en la generación  $t$ ;

y, por último,

$$P'_{j \rightarrow i}(t) = (j \rightarrow i) P'_j(t)$$

es la fracción de la población que, en la generación  $t+1$ , ocuparán los descendientes de individuos con genotipo  $j$  seleccionados en la generación  $t$  que mutan en el genotipo  $i$ .

Entonces la ecuación (1.1) se puede escribir como

$$P_i(t+1) = \sum_{j=1}^{2^N} P'_{j \rightarrow i}(t) \tag{1.2}$$

y leerse de la siguiente manera: “La fracción de la población de genotipos  $i$  que se tendrá en la generación  $t+1$  será igual a la suma de todas las fracciones de la población que ocuparán los descendientes de individuos  $j$  seleccionados en la generación  $t$  que mutan en el genotipo  $i$ ”.

En el resto del capítulo y de esta tesis se hará uso de esta ecuación de forma analítica y numérica (junto con otras herramientas) para explicar cada situación iniciando desde las más simples haciéndose cada vez más complejas.

## 1.2 Casos extremos

Para iniciar el estudio, es necesario partir de lo más simple para luego introducir más elementos que hacen al sistema más complejo. A continuación se analizan los casos en donde los operadores genéticos tienen valores extremos en los que, o no tienen ningún efecto o tienen un efecto extremo.

Se dice que el sistema está en un **punto de equilibrio** cuando se cumple

$$P_i(t+1) = P_i(t) \quad \forall t \geq t_0 \tag{1.3}$$

Esto significa que el efecto de la selección se ve compensado por el efecto de la mutación y las concentraciones de todos los genotipos se mantienen constantes.

### 1.2.1 Operadores con efecto nulo

En un sistema genético que evoluciona bajo la acción de la selección y la mutación, hay muchos puntos de equilibrio posibles. Uno de ellos es cuando los operadores genéticos simplemente no actúan quedándose la población exactamente como estaba inicialmente (descartando la deriva genética que no está presente en el modelo con población infinita). Este no es el único punto de equilibrio posible. Hay otros puntos de equilibrio que son consecuencia de situaciones extremas en los parámetros del sistema y se describen a continuación.

### 1.2.2 Cuando la probabilidad de mutación es 1/2

Si en la ecuación (1.1) se sustituye  $p = 1/2$  se obtiene:

$$P_i(t+1) = \sum_{j=1}^{2^N} \frac{(1/2)^{2_{ij}} (1/2)^{N-2_{ij}}}{f(t)} f_j P_j(t) = \frac{(1/2)^{2^N}}{f(t)} \sum_{j=1}^{2^N} f_j P_j(t) = \frac{(1/2)^{2^N}}{f(t)} \bar{f}(t) = \frac{1}{2^N},$$

es decir, que todos los genotipos ocuparán la misma fracción de población después de una sola generación obteniéndose una **distribución de población uniforme** y en **equilibrio** sin importar cuál sea el *paisaje de aptitud* ni la *población inicial*.

### 1.2.3 Población uniformemente distribuida y en equilibrio

¿Cuál es la condición para que una población uniformemente distribuida esté en equilibrio? De la ecuación (1.1) se tiene que, con una **población uniforme** y en **equilibrio** se tendría

$$\frac{1}{2^N} = \sum_{j=1}^{2^N} p^{2_{ij}} (1-p)^{N-2_{ij}} \frac{f_j}{f(t)} \frac{1}{2^N}$$

para toda  $i$ . De aquí se obtiene

$$\bar{f}(t) = \prod_{j=1}^{2^N} p^{2_{ij}} (1-p)^{N-2_{ij}} f_j$$

$$\frac{1}{2^N} \prod_{k=1}^{2^N} f_k = \prod_{j=1}^{2^N} p^{2_{ij}} (1-p)^{N-2_{ij}} f_j$$

Resolviendo esta última ecuación para  $N = 1, 2, 3, 4$  con el paquete simbólico comercial **Maple7** [ 39 ] se obtienen solamente 2 posibles soluciones:

- ?  $p = 1/2$  sin restricciones en las  $f_i$  y
- ?  $f_1 = f_2 = f_3 = \dots = f_{2^N}$ , es decir, un **paisaje de aptitud plano** sin restricciones para  $p$ .

Se puede demostrar que éstas son las únicas soluciones para toda  $N$  de la siguiente manera: La primera solución se obtiene como en la sección 1.2.2 y la segunda es el caso cuando ninguna secuencia tiene ventaja reproductiva sobre las demás. Si no fuera ninguno de estos dos casos, habría al menos una secuencia con mayor aptitud y, consecuentemente, con ventaja reproductiva, y se tendría alguno de dos casos:  $p < 1/2$  ó  $p > 1/2$ . El primero implica que las secuencias genéticas de los individuos tienden más a conservar sus genes que a modificarlos y la segunda lo contrario. Cuando  $p < 1/2$ , la ventaja reproductiva de la secuencia con mayor aptitud que las otras no quedaría completamente eliminada por mutaciones (ya que hay una tendencia a conservar sus genes) quedando un remanente que rompería la condición de equilibrio, es decir, la distribución de la población ya no sería la misma que en la generación anterior. Cuando  $p > 1/2$ , dicha ventaja se transformaría en ventaja para la secuencia opuesta (y sus vecinas) rompiéndose igualmente la condición de equilibrio.

### 1.2.4 Paisaje de aptitud plano y población en equilibrio

Si no hay selección ¿Cuándo puede haber equilibrio? Con un paisaje de aptitud plano en equilibrio se tiene de (1.1)

$$P_i(t) = \prod_{j=1}^{2^N} p^{2_{ij}} (1-p)^{N-2_{ij}} \frac{f_0}{\bar{f}(t)} P_j(t)$$

y

$$P_i(t) = \prod_{j=1}^{2^N} p^{2_{ij}} (1-p)^{N-2_{ij}} P_j(t)$$

Resolviendo con **Maple** [ 39 ] para  $N = 1, 2, 3$  se obtienen 3 posibles soluciones que son:

- ?  $p = 0$  sin restricciones para las  $P_i(t)$ , (acción nula de los operadores)
- ? población con *distribución uniforme* sin restricciones para  $p$  y
- ?  $p = 1$  con las restricciones  $P_i(t) = P_{\bar{i}}(t)$  donde  $\bar{i}$  es el complemento de  $i$ , es decir, la **secuencia opuesta** de la secuencia  $i$ . A esta distribución se le llamará **distribución simétrica**.

Para demostrar que no hay más soluciones para toda  $N$  se puede razonar así: Supóngase que  $p = 1$  y que la distribución de la población no es simétrica. En este caso todos los individuos con secuencia genética  $i$ , mutan hacia la secuencia  $\bar{i}$  y viceversa. Esto implica que las cantidades  $P_i$  y  $P_{\bar{i}}$  simplemente intercambian sus valores en la siguiente generación y por lo tanto deben ser iguales para mantener la condición de equilibrio, es decir, la población debe ser simétrica. De no ser simétrica, se tiene un equilibrio cíclico de dos estados. Si  $p < 1$  y la población no es uniforme, entonces, aunque ninguna secuencia tiene ventaja reproductiva, habrá un desequilibrio causado por la mutación, ya que, al haber más individuos de alguna secuencia, también habrá más mutantes provenientes de tal secuencia y, si  $p \neq 1/2$ , éstos se distribuirán de manera no uniforme en el espacio de los genomas favoreciendo a algunas secuencias más que a otras rompiendo así la condición de equilibrio.

### 1.2.5 Solución de la ecuación de evolución para un gen ( $N = 1$ )

La ecuación (1.1) representa a un sistema de  $2^N$  ecuaciones de diferencias con  $2^N$  incógnitas que son las funciones  $P_i(t)$ . Su solución consiste en hallar las expresiones para cada una de las  $P_i(t)$  bajo cualquier paisaje de aptitud y condición inicial. Obtener estas soluciones de manera analítica es una tarea muy difícil. Para  $N = 2$  las soluciones tienen expresiones ya bastante grandes de las que difícilmente se puede obtener más información que explique el comportamiento complejo del sistema y resultan definitivamente imprácticas. Sin embargo algo puede aprenderse de la solución para un gen, que no es tan complicada, y lo aprendido puede aplicarse, no sin ciertas modificaciones, a sistemas con  $N$  mayor a 1.

El sistema genético con selección y mutación cumple con

$$x_i(t+1) = \sum_{j=1}^{2^N} (j \text{ mut } i) f_j x_j(t) \tag{1.4}$$

donde  $x_i(t)$  es la cantidad de individuos cuyo genotipo es  $i$  presentes en la población en la generación  $t$ . Si se normalizan estas  $x_i$  mediante

$$P_i(t) = \frac{x_i(t)}{\sum_j x_j(t)}$$

se llega a la ecuación (1.1). Así que las soluciones de (1.4) normalizadas serán equivalentes a las soluciones de (1.1).

Si se consideran solo genotipos de longitud uno ( $N = 1$ ), se tendrá un sistema con solo 2 posibles genotipos distintos. La solución de (1.4) para  $N = 1$  es (tomada de [ 61 ])

$$\begin{aligned} x_0(t) &= A(t) p f_1 x_1(0) + (1 - p) f_1 x_0(0) \\ x_1(t) &= A(t) p f_0 x_0(0) + (1 - p) f_0 x_1(0) \end{aligned} \quad (1.5)$$

donde

$$A(t) = \frac{e^{-\lambda t} - e^{-\mu t}}{\lambda - \mu}$$

y

$$\lambda = \frac{1}{2} (1 - p)(f_1 + f_0) + \sqrt{(1 - p)^2 (f_1 - f_0)^2 + 4(1 - 2p)f_1 f_0}$$

Pero ¿qué implica esta solución? Tras el siguiente análisis, puede llegarse a conclusiones interesantes aplicables también para  $N > 1$ .

Se demostrará a continuación que el efecto de la población inicial es despreciable cuando  $t \gg 1$  y además que, si se normalizan las variables  $x_i$ , las nuevas variables normalizadas tienden a un valor constante cuando  $t \gg 1$ . Luego se analiza la expresión para este valor constante y se le da una interpretación en términos de la evolución genética

**Suposición 1:**  $|f_1| < |f_0|$ . (1.6)

$A(t)$  puede escribirse como

$$A(t) = \frac{e^{-\lambda t} - e^{-\mu t}}{\lambda - \mu} = \frac{e^{-\lambda t} (1 - e^{-(\mu - \lambda)t})}{\lambda - \mu}$$

entonces, cuando  $t \gg 1$

$$A(t \gg 1) \approx \frac{e^{-\lambda t}}{\lambda - \mu}$$

debido a la suposición 1, por lo tanto se tiene que

$$\begin{aligned} x_0(t \gg 1) &\approx \frac{e^{-\lambda t}}{\lambda - \mu} p f_1 x_1(0) + (1 - p) f_1 x_0(0) \\ x_1(t \gg 1) &\approx \frac{e^{-\lambda t}}{\lambda - \mu} p f_0 x_0(0) + (1 - p) f_0 x_1(0) \end{aligned}$$

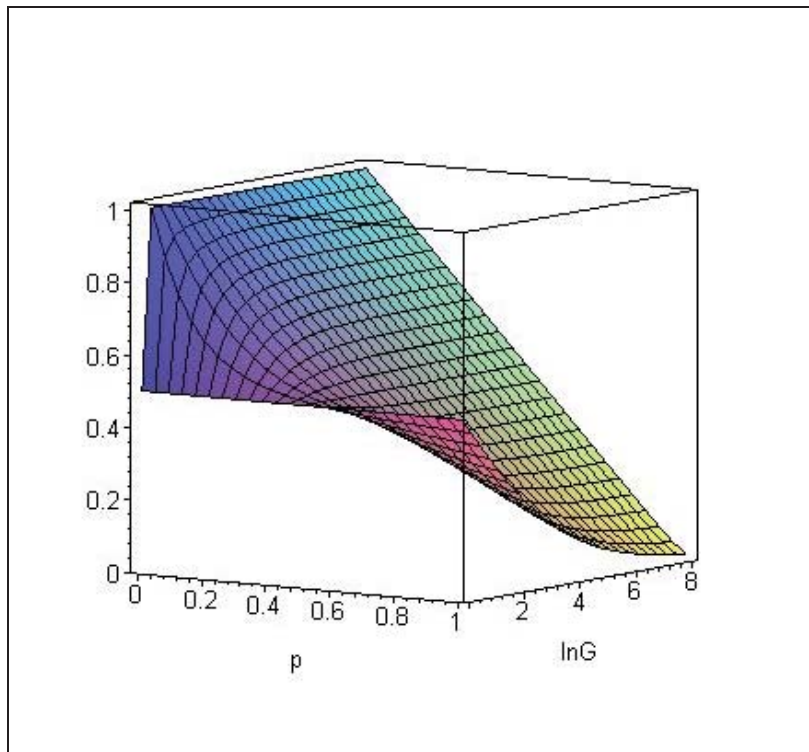
y entonces, al normalizar, se cancela  $A(t)$  y también  $x_0(0)$  y  $x_1(0)$

$$\begin{aligned}
 P_0(t \gg 1) &\approx \frac{x_0(t)}{x_0(0) + x_1(t)} \\
 &\approx \frac{A(t) + pf_1 x_1(0) + (1-p)f_1 x_0(0) + \gamma x_0(0)}{A(t) + pf_1 x_1(0) + (1-p)f_1 x_0(0) + \gamma x_0(0) + pf_0 x_0(0) + (1-p)f_0 x_1(0) + \gamma x_1(0)} \\
 &\approx \frac{pf_1 x_1(0) + (1-p)f_1 + \gamma x_0(0)}{\gamma pf_0 + (1-p)f_1 + \gamma x_0(0) + pf_1 + (1-p)f_0 + \gamma x_1(0)} \\
 &\approx \frac{pf_1}{pf_1 + (1-p)f_0 + \gamma} \approx K_0
 \end{aligned}$$

que depende solamente de  $p, f_0$  y  $f_1$ .

De la misma manera se llega a

$$P_1(t \gg 1) \approx \frac{pf_0}{pf_0 + (1-p)f_1 + \gamma} \approx K_1 \approx 1 - K_0 \tag{1.7}$$



**Figura 1-1** El valor de  $K_1$  como función de  $p$  y de  $\ln(G) = \ln(f_1/f_0)$ .

En la **Figura 1-1** se presenta la gráfica de  $K_1$  como función de  $p$  y de  $\ln(G)$  donde  $G = f_1/f_0$  se define como la ganancia de aptitud. Nótese que, para valores grandes de  $G$ ,  $K_1$  depende casi



linealmente de  $p$  y para valores bajos la relación es no lineal. Esto se debe a que

$$\lim_{f_1 \rightarrow 1} K_1 \approx 1 - p$$

que es la probabilidad de que el bit que conforma a cada genotipo no mute.

Esto significa que, cuando uno de los genotipos domina mucho al otro ( $G$  grande), la concentración del segundo depende casi exclusivamente de la cantidad de mutantes provenientes del que domina, lo cual está determinado solamente por la tasa de mutación  $p$ . El genotipo de mayor aptitud domina totalmente cuando  $p = 0$  ya que no se generan mutantes y, con  $p$  cercana a 1, el dominio es casi total pero del otro genotipo ya que casi todos los genotipos mutan hacia el otro.

Cuando  $G$  es grande, se dice que la selección es fuerte ya que solamente se selecciona a los de mayor aptitud (aunque éstos luego muten) eliminándose rápidamente a todos los que tienen aptitud (mucho) menor. El GLE en este caso es simplemente la tasa de mutación la cual distribuye la población, en cada generación, siempre desde el genotipo dominante.

En el caso con selección débil, cuando  $G$  no es muy grande, algunos de los individuos con genotipo no dominante quedan en la población para la siguiente generación. En este caso hay más GLEs que la tasa de mutación interviniendo el valor de  $G$  también como un GLE. Por lo tanto el sistema es más complejo cuando la selección es débil que cuando es muy fuerte.

El resultado dado en (1.7) se basa en la suposición 1 (1.6). Las únicas dos maneras de que no se cumpla (1.6) son cuando  $p = 0$  y  $f_0 = f_1$  y cuando  $p = 1$  que son casos ya estudiados en la sección 1.2. En estos casos se cumple:

$$|f_1 / f_0| = 1 \tag{1.8}$$

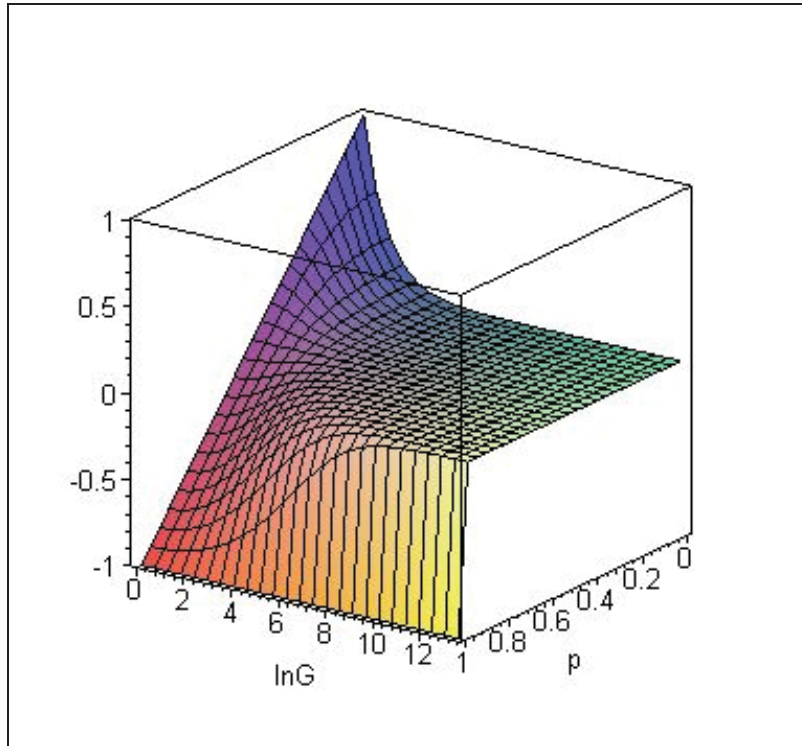
y ningún otro caso es posible.

En la **Figura 1-2** se dibuja el cociente  $f_1 / f_0$  como función de  $p$  y de  $\ln(f_1/f_0)$ . Cuando el valor del cociente  $f_1 / f_0$  es cercano a cero, el efecto transitorio de la población inicial será despreciable en pocas generaciones y cuando dicho valor sea cercano a 1 o  $-1$  el efecto será importante durante un número grande de generaciones. Y el efecto es permanente cuando  $|f_1 / f_0| = 1$ . El efecto transitorio es importante, entonces, cuando hay selección débil y la tasa de mutación es lejana a  $1/2$ .

Puede verse de la expresión de  $f_1$  que se cumplen  $f_1 \geq 1$  y  $f_1 \leq f_0 \leq 1$ . Esta última será igual a cero cuando  $p = 1/2$ , mayor a cero cuando  $p < 1/2$  y menor a cero cuando  $p > 1/2$ . El cociente  $f_1 / f_0$  tendrá entonces el mismo signo que  $f_1$ .

En el caso cuando  $p > 1/2$  el factor  $A(t)$  presentará oscilaciones al paso de las generaciones ya

que  $f_1/f_0$  será negativo. La amplitud de estas oscilaciones irá disminuyendo al paso de las generaciones siempre que se cumpla (1.6) hasta llegar a ser despreciables. El efecto *no* despreciable de dichas oscilaciones tendrá una duración mayor cuando  $f_1/f_0$  esté más cerca de cumplir con (1.8), es decir, cuando sea cercano a  $-1$  (ver **Figura 1-2**). Y esto se cumple cuando  $p$  es cercana a 1 y  $\ln(G)$  es cercano a cero, es decir, cuando  $f_1$  es cercana a  $f_0$  lo que significa selección débil.



**Figura 1-2** El cociente  $f_1/f_0$  como función de  $p$  y de  $\ln(G) = \ln(f_1/f_0)$ . La duración del efecto transitorio de la población inicial depende del valor absoluto de este cociente.

En el caso  $p = 0$  y  $f_0 = f_1$  no se tiene mutación y, por lo tanto, no es posible que una secuencia se convierta en ninguna otra y, como tampoco hay diferencia en la aptitud reproductiva de las secuencias, las proporciones de la población total que ocupan todas las secuencias genéticas serán siempre las mismas. Es decir, que las variables normalizadas tendrán siempre los mismos valores que en la población inicial. Se deduce que el efecto de la población inicial es permanente.

Cuando  $p = 1$  simplemente todas las secuencias “0” mutan hacia la secuencia “1” y viceversa. Esto hace que si, al aplicar la selección, una secuencia es más “popular” que la otra, resultará exactamente lo contrario después de aplicar la mutación. En otras palabras, lo que gana la secuencia más apta, en cuanto a la fracción de la población que ocupa, es entregado a la otra secuencia en cada generación. En este caso se tiene de la ecuación (1.1)

$$\begin{aligned}
 P_0(t+1) &= \frac{f_1}{f(t)} P_1(t) = \frac{f_1 P_1(t)}{f_0 P_0(t) + f_1 P_1(t)} = \frac{f_1 P_1(t)}{f_0 P_0(t) + f_1 P_1(t)} \\
 P_0(t+2) &= \frac{f_1 P_1(t+1)}{f_0 P_0(t+1) + f_1 P_1(t+1)} \\
 &= \frac{f_1 \frac{f_1 P_1(t)}{f_0 P_0(t) + f_1 P_1(t)}}{f_0 \frac{f_0 P_0(t) + f_1 P_1(t)}{f_0 P_0(t) + f_1 P_1(t)} + f_1 \frac{f_1 P_1(t)}{f_0 P_0(t) + f_1 P_1(t)}} \\
 &= \frac{f_1 P_1(t)}{f_0 P_0(t) + f_1 P_1(t)} = P_0(t)
 \end{aligned}$$

lo que significa que cada 2 generaciones se tendrán las mismas concentraciones de cada genotipo que se tenían inicialmente, es decir, que se entra en un ciclo de dos estados. Por lo tanto, el efecto de la población inicial en este caso es también permanente en el proceso de evolución.

Los Grados de Libertad Efectivos son en este caso la fuerza de la selección  $G$  y la tasa de mutación  $p$ . Los valores de  $G$  y  $p$  determinan el estado final de la población y con  $G$  grande, el cálculo de la distribución final se simplifica. El tiempo que la población inicial tendrá un efecto significativo en las concentraciones de cada genotipo depende también de  $G$  y  $p$  y con  $G$  grande  $p$  no influye a menos que  $p$  sea muy cercana a 1. Otro GLE es el hecho de que  $p$  sea mayor a  $1/2$  o no ya que esto determina si en la fase transitoria se presentarán oscilaciones o no en las concentraciones de los genotipos.

Dado que la normalización de las variables  $x_i(t)$  en (1.4) lleva a la ecuación (1.1) las conclusiones a las que se llegó en esta sección aplican también para tal ecuación cuando  $N = 1$ .

En las siguientes secciones se revisarán las conclusiones antes descritas para compararlas con las de sistemas donde  $N > 1$  y se generalizarán para toda  $N$ .

### 1.2.6 Solución general de la ecuación de evolución

La ecuación (1.1) puede resolverse para las  $P_i(t)$  numéricamente simplemente iterando desde  $t = 0$  hasta un número fijo de generaciones. Haciendo esto pueden observarse algunas características de la evolución del sistema al variar los parámetros que aparecen en la ecuación que son:  $p$ ,  $N$ , el paisaje de aptitud y la población inicial aunque esta última en general no afecta el estado final o de equilibrio de la población después de algunas generaciones. En las siguientes secciones se presentan algunas gráficas de la dinámica del sistema obtenidas de esta manera.

### 1.3 La Cuasiespecie en el paisaje de aguja en un pajar

Eigen [ 19 ] describe lo que él llama la emergencia de una Cuasiespecie de individuos que no comparten exactamente la misma información genética sino que, en el espacio de los genomas, se agrupan alrededor de un genotipo cuya aptitud es superior a la de los demás. Cuando la población llega a un punto de equilibrio, la distribución de los individuos en el espacio de los genomas describe esta cuasiespecie. La forma de esta distribución depende de la tasa de mutación y del paisaje de aptitud. Ahora se verá cómo es esta dependencia y de qué dependen los fenómenos que emergen como la tasa de mutación crítica.

Teniendo ya un punto de referencia de los casos extremos y de la solución de la ecuación de evolución para un bit ( $N = 1$ ) se estudiará ahora la dinámica en sistemas genéticos en donde  $N$  es relativamente grande pero con un paisaje de aptitud ligeramente más complejo. Este paisaje es conocido como de “aguja en un pajar”.

El paisaje de aptitud de “aguja en un pajar” es uno en el que todos los genotipos posibles tienen el mismo fenotipo excepto uno, el cual tiene un fenotipo que lo hace superior a los demás. Este genotipo especial juega el papel de la aguja en el pajar y se le llama también secuencia genética maestra o simplemente secuencia maestra. Se le llama “aguja en un pajar” porque, en poblaciones finitas, es posible que no se tenga a la aguja o secuencia maestra en la población y, en este caso, la población no es llevada hacia la aguja mediante mutaciones en el proceso evolutivo ya que el resto del paisaje es plano y, por lo tanto, ninguna secuencia es preferida sobre otra, lo que hace que la búsqueda de la secuencia maestra sea muy difícil. Existen variaciones de este paisaje en las que no solamente es una la secuencia maestra sino varias (degeneración) que se estudian más adelante. Dichas secuencias maestras pueden estar todas “juntas” la una a la otra (a una sola mutación de separación) o en grupos dispersos en el dominio de los genotipos. Estas variaciones se estudian más adelante también.

En el caso particular del paisaje de aptitud de “aguja en un pajar” se deben definir dos valores, uno para la aptitud de la secuencia maestra y otro para la de las demás secuencias.

$f_{sm}$  = aptitud de la secuencia maestra.

$f_0$  = aptitud de las demás secuencias.

Se definirá también el concepto de *ganancia de aptitud*  $G$  como

$$G = \frac{f_{sm}}{f_0}.$$

A la  $P_i(t)$  que corresponde a la secuencia maestra se le denominará  $P_{sm}(t)$ . Así que, usando la ecuación (1.1), se tiene que, para población infinita,

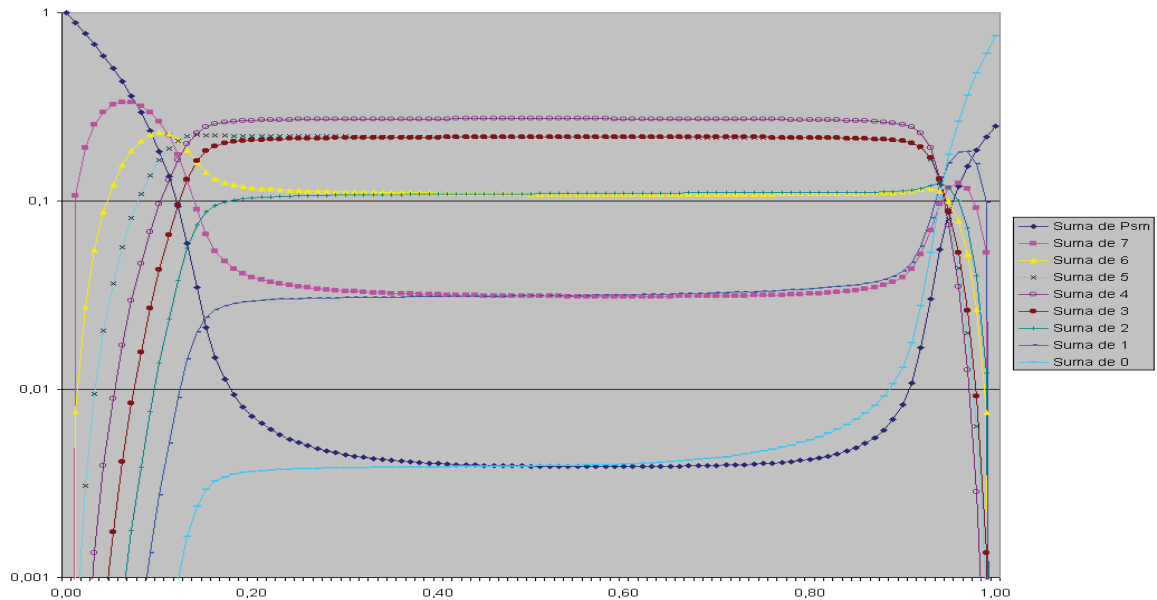
$$P_{sm}(t) = (1-p)^N \frac{f_{sm}}{f(t)} P_{sm}(t) + \sum_{j \neq sm} p^{sm,j} (1-p)^{N-1} \frac{f_0}{f(t)} P_j(t) \quad (1.9)$$

El primer término del lado derecho de esta ecuación (término de conservación) es la fracción de la población de *secuencias maestras seleccionadas* que no sufren mutaciones y el segundo término (término de reconstrucción) es la *fracción de la población seleccionada* distinta a la secuencia maestra que muta en la secuencia maestra.

Para las demás secuencias la ecuación es similar a la de la secuencia maestra. En estos casos, a pesar de tener el mismo valor de aptitud, las ecuaciones no son idénticas (al menos no todas). Esta diferenciación se conoce como un rompimiento de la simetría del paisaje de aptitud inducido por el efecto de la mutación. Esto hace que las concentraciones para las secuencias no maestras sean diferentes entre sí dando lugar a la formación de una Cuasiespecie.

### 1.3.1 La “forma” de la Cuasiespecie

Como ya se dijo, la Cuasiespecie es una población cuya genética no es exactamente igual pero que está centrada alrededor de una secuencia dominante que atrae hacia ella la genética del conjunto de la población. Surge la pregunta de ¿Cuál es la distribución de concentraciones de cada una de las secuencias cercanas y no tan cercanas a la secuencia maestra?

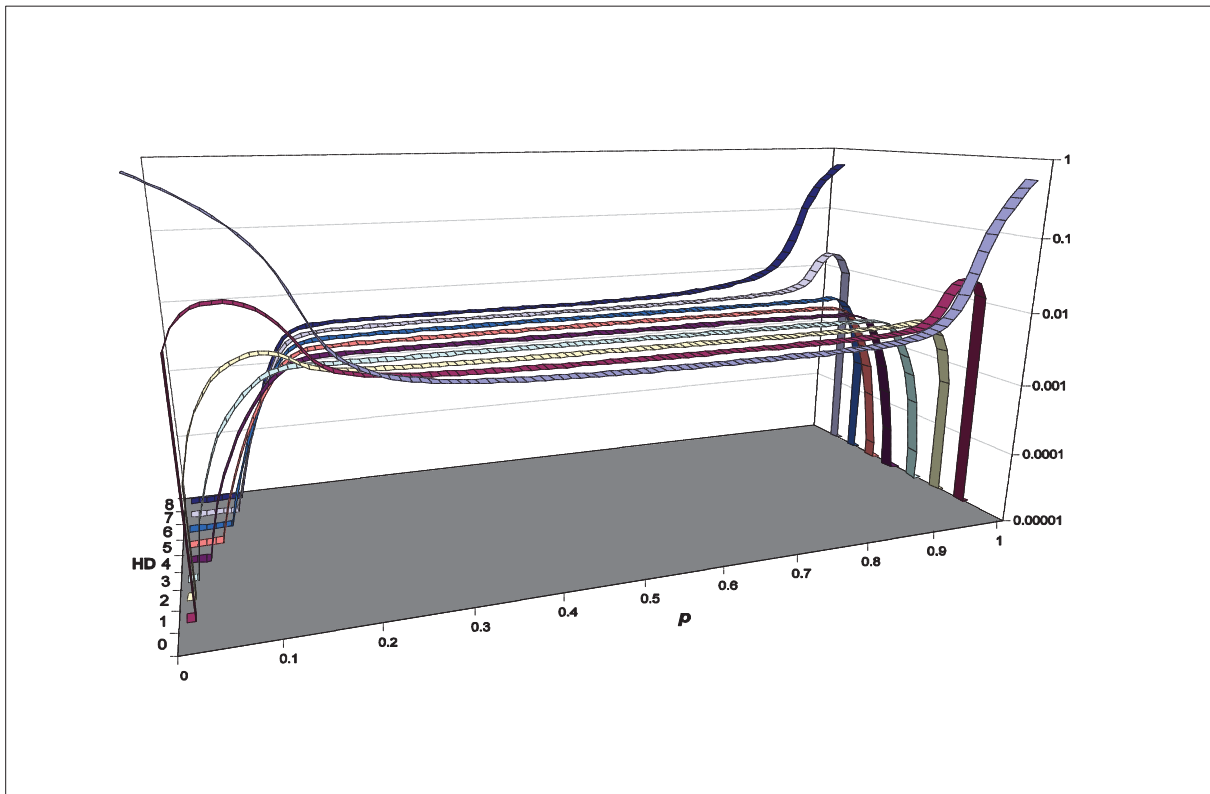


**Figura 1-3** Concentraciones como función de  $p$  después de 50 generaciones a una  $G = 3$  y población inicial uniforme de 8 genes. Las leyendas Suma de  $h$  indican al conjunto de secuencias con  $h$  genes en 1 o la clase de Hamming  $h$ .

En la **Figura 1-3** se presenta una gráfica de los resultados de la integración de la ecuación (1.1) en el paisaje de aguja en un pajar con  $G = 3$ . Se utilizó una población inicial (infinita) uniforme de individuos cuyos genotipos son secuencias de 8 genes y la secuencia maestra es la que tiene 8 genes en uno. Se muestran las concentraciones, después de 50 generaciones (ya prácticamente en estado estacionario), de cada *clase de Hamming*  $h$ , es decir, la suma de las concentraciones de las  ${}^N C_h$  secuencias que tienen  $h$  genes en 1 (coincidentes con la secuencia

maestra) y los demás en 0. La clase 8 corresponde a la secuencia maestra.

Lo primero que se nota en la **Figura 1-3** es que cada clase de Hamming diferente a la clase 8 tiene una respuesta diferente a pesar de que todas tienen el mismo fenotipo, es decir, que la acción conjunta de la selección con la mutación ha inducido un **rompimiento de la simetría** del paisaje, en donde se tienen solamente 2 fenotipos y, sin embargo, hay más de dos comportamientos distintos.



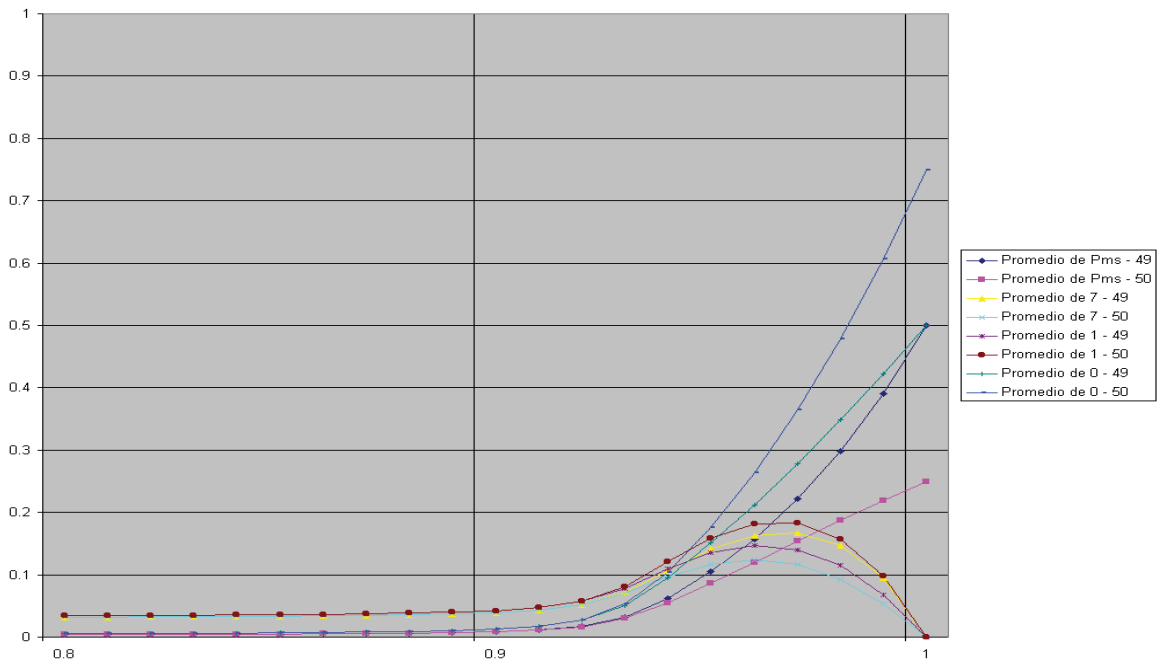
**Figura 1-4** Concentraciones normalizadas como función de  $p$  de cada clase de Hamming (HD por Hamming Distance) después de 50 generaciones a una  $G = 3$  y  $N = 8$ .

Para distinguir el efecto de la mutación en cada secuencia perteneciente a una misma clase de Hamming, se muestran en la **Figura 1-4** las mismas concentraciones que en la **Figura 1-3** pero normalizadas por el número de secuencias genéticas que incluye cada clase para conocer la concentración de cada secuencia genética individualmente.

Para  $p < 1/2$ , las secuencias que pertenecen a las clases de Hamming más alejadas de la secuencia maestra, tienen menor concentración y las más cercanas mayor, formándose lo que en [ 19 ] se llama una Cuasiespecie centrada alrededor de la secuencia maestra. Los individuos de esta Cuasiespecie tienen genética similar a la de la secuencia maestra.

Para  $p > 1/2$  la situación es diferente ya que ahora son favorecidas las secuencias cercanas a la secuencia maestra pero también las más lejanas que son cercanas a la secuencia opuesta de la maestra. Esto es especialmente notorio cuando  $p$  es cercana a 1. Entonces la Cuasiespecie está

dividida en dos aglomeraciones, una alrededor de la secuencia maestra y la otra alrededor de su opuesta.



**Figura 1-5** Gráfica de  $P_{sm}$  y de las clases de Hamming 0, 1, y 7 como función de  $p$  cerca de  $p = 1$  en las generaciones 49 y 50 a partir de una población inicial uniforme de individuos con 8 genes. No se muestran las clases 2, 3, 4, 5, 6 por visibilidad pero tienen un comportamiento similar a las clases 1 y 7 cerca de  $p = 1$ .

Cuando  $p = \frac{1}{2}$  la cuasiespecie es plana ya que todos los genotipos están igualmente representados. Cuando  $p = 0$  la secuencia maestra es la única que está representada (en el punto de equilibrio) y por lo tanto la Cuasiespecie es como una espiga delgada situada sobre la secuencia maestra (en el dominio de los genotipos). Cuando  $p$  es algo intermedio entre 0 y  $\frac{1}{2}$  la Cuasiespecie toma una “forma” en la que la secuencia maestra es la que ocupa la mayor proporción de la población mientras que la proporción de las demás secuencias individuales disminuye como función de su distancia de Hamming con la secuencia maestra. En otras palabras, para valores de  $p$  cercanos a 0, la forma de la Cuasiespecie es alta y delgada ya que hay altas concentraciones de pocas secuencias, mientras que para valores de  $p$  cercanos a 0.5 la forma es baja y ancha ya que todos los genotipos tienen poca concentración.

En la **Figura 1-5** se observa que cuando  $p$  tiende a 1 hay una *bifurcación* en el valor de  $P_{sm}$  entre las generaciones 49 y 50. A continuación se da una explicación del porqué de estos resultados.

### 1.3.1.1 La Cuasiespecie cuando $p = 0$

Con  $p = 0$  en la ecuación (1.1) se tiene

$$P_i(t+1) = \frac{f_i}{\bar{f}(t)} P_i(t)$$

que en el caso de la secuencia maestra es

$$P_{sm}(t+1) = \frac{f_{sm}}{\bar{f}(t)} P_{sm}(t)$$

Considerando que  $f_{sm} > \bar{f}(t)$  resulta que  $P_{sm}(t+1) > P_{sm}(t)$  para toda  $t$ .

Para las demás secuencias  $j$  se cumple que  $f_j < \bar{f}(t)$  y por lo tanto  $P_j(t+1) < P_j(t)$ , es decir, que  $P_j(t)$  tiende a cero para valores grandes de  $t$ . Véase en la **Figura 1-3** cómo todas las concentraciones de las *clases de Hamming* de las secuencias no maestras tienden a cero cuando  $p$  tiende a 0.

Lo anterior demuestra que, para  $p = 0$ , la secuencia maestra tiende a dominar toda la población.

### 1.3.1.2 La Cuasiespecie cuando $p = 1$ :

Con  $p = 1$  en (1.1) se tiene

$$P_i(t+1) = \frac{f_{\bar{i}}}{\bar{f}(t)} P_i(t)$$

donde  $\bar{i}$  es la **secuencia opuesta** a  $i$ , es decir, aquella secuencia que difiere en todos sus genes con la secuencia  $i$ .

Para las secuencias no maestras cuya opuesta no es la secuencia maestra se tiene lo siguiente

$$P_i(t+1) = \frac{f_0}{\bar{f}(t)} P_i(t) \tag{1.10}$$

$$P_i(t+2) = \frac{f_0}{\bar{f}(t+1)} \frac{f_0}{\bar{f}(t)} P_i(t) = P_i(t)$$

ya que  $f_0 < \bar{f}(t)$ ,  $f_0 < \bar{f}(t+1)$ , es decir, que cada 2 generaciones se reduce la cantidad  $P_i$ , lo que implica que estas secuencias tienden a extinguirse.

Para la secuencia maestra y su opuesta se tiene



$$P_{sm}(t+1) = \frac{f_0}{f(t)} P_{\overline{sm}}(t)$$

$$P_{\overline{sm}}(t+1) = \frac{f_{sm}}{f(t)} P_{sm}(t)$$

y también

$$P_{sm}(t+2) = \frac{f_0}{f(t+1)} \frac{f_{sm}}{f(t)} P_{sm}(t)$$

De esto último se puede ver que el factor por el que se multiplica  $P_{sm}(t)$  cada dos generaciones es mayor al factor dado en (1.10) ya que  $f_0 > f_{sm}$ . Algo similar sucede con la secuencia opuesta. Se deduce entonces que, tanto la secuencia maestra como su opuesta, dominarán la población. Véase en la **Figura 1-5** cómo todas las *clases de Hamming* de las secuencias no maestras, excepto la clase 0 (que es la opuesta de la secuencia maestra), tienden a cero cuando  $p$  tiende a 1.

Si se define ahora la variable

$$z(t) = \frac{P_{sm}(t)}{P_{\overline{sm}}(t)}, \quad P_{\overline{sm}}(t) > 0$$

se tiene que, para  $p = 1$ ,

$$z(t+1) = \frac{P_{sm}(t+1)}{P_{\overline{sm}}(t+1)} = \frac{\frac{f_0}{f(t)} P_{\overline{sm}}(t)}{\frac{f_{sm}}{f(t)} P_{sm}(t)} = \frac{f_0}{f_{sm}} \frac{1}{z(t)} = \frac{1}{Gz(t)}$$

y también

$$z(t+2) = \frac{1}{Gz(t+1)} = \frac{1}{G \frac{1}{Gz(t)}} = z(t)$$

lo que indica que la proporción entre la secuencia maestra y su opuesta es la misma cada 2 generaciones. Es decir, la población se encuentra en una oscilación permanente de período 2.

Si, por ejemplo, la población inicial es uniforme, se cumple  $z(0) = 1$ , luego  $z(1) = 1/G$  lo que indica que, en la generación 1, hay  $G$  secuencias opuestas por cada secuencia maestra. En la siguiente generación se tendría nuevamente  $z(2) = z(0) = 1$ , y así alternadamente en cada generación. Esto significa que habrá una oscilación permanente tanto para  $P_{sm}$  como para  $P_{\overline{sm}}$ .

Esta oscilación puede notarse en la **Figura 1-5** en donde se observan bifurcaciones entre las curvas para las generaciones 49 y 50 cerca de  $p = 1$  tanto para  $P_{sm}$  como para la secuencia opuesta que es la *clase de Hamming 0*. Tales oscilaciones tienden a reducir su amplitud conforme aumenta el número de generaciones. La rapidez de esta reducción depende de qué tan cerca esté  $p$  de 1, mientras más cerca será más lenta. Cuando  $p$  está cerca de 1, hay un fuerte dominio de la secuencia maestra y de su opuesta quedando prácticamente un sistema donde solamente hay 2 secuencias que mutan la una hacia la otra (como en el caso de un gen de la sección 1.2.5).

### 1.3.2 La tasa crítica de mutación de una Cuasiespecie

Una observación importante en la **Figura 1-4** es que para un rango bastante amplio de  $p$  alrededor del valor  $\frac{1}{2}$ , se tiene que todas las secuencias tienen prácticamente la misma concentración. Hay un valor de  $p$  entre 0 y  $\frac{1}{2}$  a partir del cual hay un cambio abrupto en el comportamiento de las concentraciones en estado estacionario. Hay también otro valor de  $p$  entre  $\frac{1}{2}$  y 1 en donde hay otro cambio abrupto. Se dice que en estos puntos hay una transición de fase de segundo orden ya que el cambio es abrupto pero las curvas siguen siendo continuas. Véase también la sección 1.5.1 en la página 71.

A continuación se hará un cálculo aproximado, a partir de la ecuación de evolución, de esta tasa de mutación conocida como la tasa crítica de mutación que se denotará como  $p^*$ .

#### 1.3.2.1 Aproximaciones teóricas para la tasa crítica de mutación $p^*$

Si se supone despreciable el efecto de las mutaciones desde secuencias no maestras hacia la secuencia maestra, la ecuación (1.9) puede aproximarse como

$$P_{sm}(t) \approx (1-p)^N \frac{f_{sm}}{f(t)} P_{sm}(t) \tag{1.11}$$

Para poder hacer esta suposición debe considerarse lo siguiente.

Para  $N \gg 1$  y  $p$  pequeña, resulta muy poco probable que una secuencia no maestra  $j$  con  $\delta_{sm,j} = 1$  mute hacia la secuencia maestra ya que la única mutación viable es aquella en la que muta justamente el gen en el que difieren ambas secuencias que es uno entre  $N$  posibles. Pero ¿qué tan pequeña debe ser  $p$ ?

Comparando la relación entre la contribución del *término de conservación* de la ecuación (1.9) y la suma de las contribuciones de los  ${}^N C_1$  *términos de reconstrucción* que tienen  $\delta_{sm,j} = 1$  se debería tener un número mucho mayor que 1 para que las últimas sean despreciables, es decir,

$$\frac{(1-p)^N \frac{f_{sm}}{f(t)} P_{sm}(t)}{p(1-p)^{N+1} \frac{f_0}{f(t)} P_j(t)} \approx \frac{(1-p)^N \frac{f_{sm}}{f(t)} P_{sm}(t)}{p(1-p)^{N+1} \frac{f_0}{f(t)} P_j(t)} \approx \frac{p GP_{sm}(t)}{p NP_{j1}(t)} \gg 1$$

Aquí  $\bar{P}_{2,1}(t)$  es el promedio de las  $P_j(t)$  con  $\rho_{sm,j} = 1$ .

Considerando que

$$P_{sm}(t) > \bar{P}_{2,1}(t) \tag{1.12}$$

se tiene

$$\frac{p}{N} \frac{G}{\bar{P}_{2,1}(t)} > \frac{p}{N} \frac{G}{P_{sm}(t)} \gg 1,$$

es decir,

$$p \gg \frac{G}{N} \tag{1.13}$$

Haciendo lo mismo para los  ${}^N C_2$  términos que tienen  $\rho_{sm,j} = 2$  se tendrá

$$\frac{p^2}{N(N-1)} \frac{G}{\bar{P}_{2,1}(t)} > \frac{p^2}{N(N-1)} \frac{2G}{P_{sm}(t)} > \frac{p^2}{N(N-1)} \frac{2}{p} \frac{G}{N} \gg 1$$

lo cual se cumple siempre que (1.13) es verdadera y además

$$\frac{p}{N-1} > 1,$$

es decir,

$$p > \frac{2}{N-1} \tag{1.14}$$

Comparando (1.13) con (1.14) se observa que (1.13) será menos restrictiva que (1.14) cuando

$$G \gg \frac{2N}{N-1} > 2 \tag{1.15}$$

En resumen, para poder hacer la aproximación dada en (1.11), cuando se cumple (1.15) debe cumplirse (1.14) y cuando no se cumple (1.15) debe cumplirse (1.13). Se tiene entonces que (1.14) es una *cota superior* para  $p$  pero esta cota podría ser todavía menor si no se cumple con (1.15).

Para  $\beta_{sm,j} = 3$  se obtiene

$$\frac{1-p}{p} \frac{G}{N C_3} \frac{1-p}{p} \frac{1-p}{p} \frac{6G}{N(N-1)(N-2)}$$

$$\frac{1-p}{p} \frac{1-p}{N-2} \frac{1-p}{N-1} \frac{1-p}{p} \frac{G}{N} \gg 1$$

lo que se cumple siempre que (1.13) y (1.14) sean verdaderas y

$$\frac{1-p}{p} \frac{3}{N-2} > 1,$$

es decir,

$$p < \frac{3}{N-1} \tag{1.16}$$

que es menos restrictiva que (1.14). De la misma manera se puede demostrar que para  $\beta_{sm,j} > 3$  se llega a relaciones cada vez menos restrictivas. Por lo tanto, basta con las conclusiones antes descritas para poder hacer la aproximación dada en (1.11).

Ahora, por otro lado, en el punto de equilibrio (1.11) se convierte en

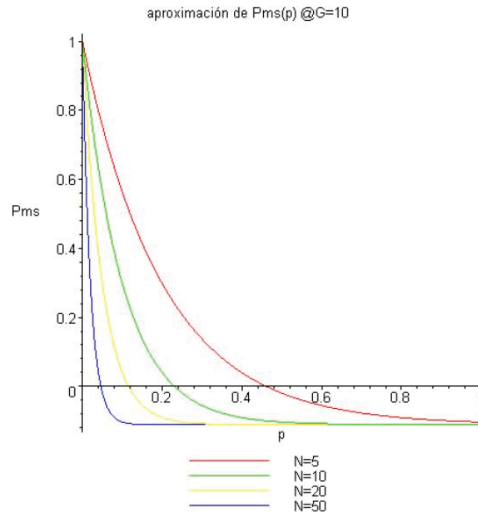
$$\begin{aligned} (1-p)^N \frac{f_{sm}}{f(t)} P_{sm} &= P_{sm} \\ \overline{f(t)} &= (1-p)^N f_{sm} \\ f_{sm} P_{sm} &= f_0(1-p) (1-p)^N f_{sm} \end{aligned}$$

Usando la definición de **Ganancia de aptitud** se tendría que, al despejar  $p$ ,

$$\begin{aligned} p &= 1 - \left( \frac{(G-1)P_{sm}}{G} \right)^{1/N} = 1 - e^{\frac{1}{N} \ln \left( \frac{(G-1)P_{sm}}{G} \right)} \\ p &= 1 - \frac{1}{N} \ln \left( \frac{G}{(G-1)P_{sm}} \right) = \frac{1}{N} \ln \left( \frac{G}{(G-1)P_{sm}} \right) \end{aligned} \tag{1.17}$$

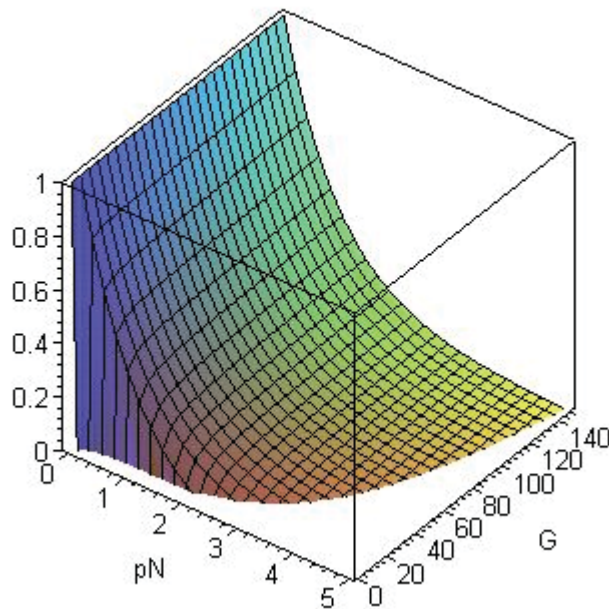
Si  $(G-1)P_{sm} \ll 1$  entonces

$$p \approx \frac{1}{N} \ln G \tag{1.18}$$



**Figura 1-6** Aproximación de  $P_{sm}$  cuando  $G = 10$  para varios valores de  $N$ . Los valores negativos se dan porque la aproximación ya no es buena en tales casos.

Aproximación de  $P_{sm}$



**Figura 1-7** Aproximación de  $P_{sm}$  como función de  $G$  y el producto  $pN$  cuando se cumplen las restricciones para  $p$  dadas en esta sección.

La condición de que  $(G-1)P_{sm} \ll 1$  se consigue “fácilmente” cuando se tiene una población uniforme ya que  $P_{sm}$  debe ser  $1/2^N$  y se debe cumplir entonces  $G \ll 2^N+1$  y, como  $N \gg 1$ , se tiene un rango amplio para  $G$  en el que se puede tener tal condición. Por lo tanto, (1.18) es una buena aproximación de  $p$  en este caso (población uniforme). A este valor de  $p$  se le llama la **probabilidad de mutación crítica**, ya que, cuando  $p$  es mayor o igual a este valor se tendrá el

mismo número de secuencias maestras que de cualquier otra secuencia no maestra. En otras palabras, sería *igualmente probable* que existiera la secuencia maestra que cualquiera de las demás secuencias. Esto significa que, en un sistema genético real, la secuencia maestra tendería a extinguirse en lugar de dominar a la población.

Ahora bien, de (1.17) se obtiene

$$P_{sm} \approx \frac{Ge^{2pN} - 1}{G - 1}, \quad (1.19)$$

que es el nivel de  $P_{sm}$  esperado dadas  $G$ ,  $p$  y  $N$ .

En la **Figura 1-6** se muestra la gráfica de la aproximación (1.19) como función de  $p$  considerando  $G = 10$  para diferentes valores de  $N$ . Si se escalara el eje de la  $p$  por  $N$  se observaría el mismo comportamiento para  $N$  grandes, es decir, que la  $P_{sm}$  depende cualitativamente del producto  $pN$ .

En la **Figura 1-7** se muestra a  $P_{sm}$  como función del producto  $pN$  y de  $G$ . Nótese que para valores muy pequeños de  $P_{sm}$  (que es el caso cuando la población es uniforme y  $N$  es grande) la relación entre  $pN$  y  $G$  es aproximadamente logarítmica, como se indica en (1.18).

### 1.3.2.2 Tasa de mutación crítica y Aptitud Efectiva de la secuencia maestra

La existencia de una tasa de mutación crítica puede explicarse mediante el concepto de *Aptitud Efectiva*.

**Definición:** La *aptitud efectiva* de un genotipo  $i$  es el nivel de aptitud de  $i$  que en caso de aplicarse solamente la selección sería necesario para obtener la misma dinámica que la que se obtiene al aplicar múltiples operadores y está dada por

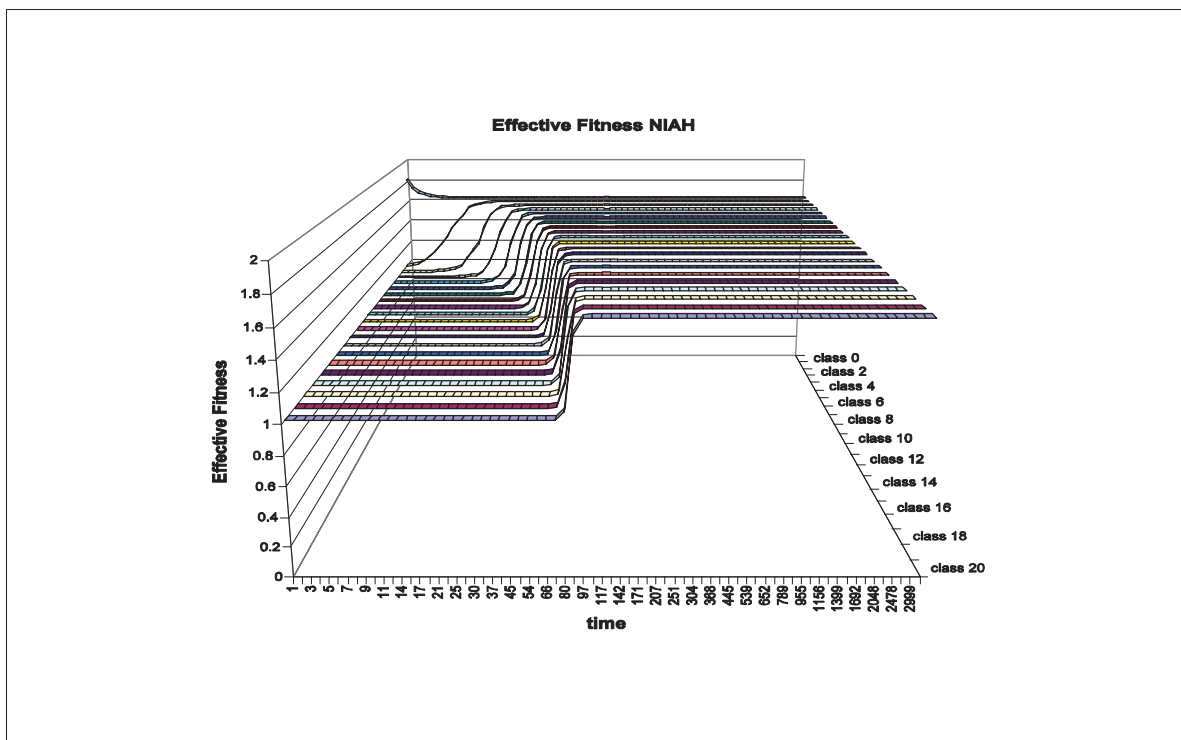
$$f_{eff}^i(t) \approx \frac{P_i(t) - 1}{P_i(t)} \bar{f}(t)$$

donde  $\bar{f}(t)$  es la aptitud promedio de la población al tiempo  $t$ .

En el presente caso se tiene la acción de la mutación la cual afecta el valor de la aptitud efectiva de la secuencia maestra (y de las demás) ya que algunos individuos con esta secuencia sufrirán mutaciones mientras que otros con secuencias distintas mutarán hacia la secuencia maestra. Este flujo de mutantes es desigual siempre que haya más individuos de alguna secuencia que de otra modificando la aptitud efectiva de cada secuencia en distinta forma.

En la **Figura 1-8** se muestra la aptitud efectiva en función del tiempo para una población inicial infinita uniforme con una tasa de mutación de 0.01 que es mucho menor a la tasa crítica de la aguja 0.034. Se muestran las distintas clases de Hamming respecto a la secuencia

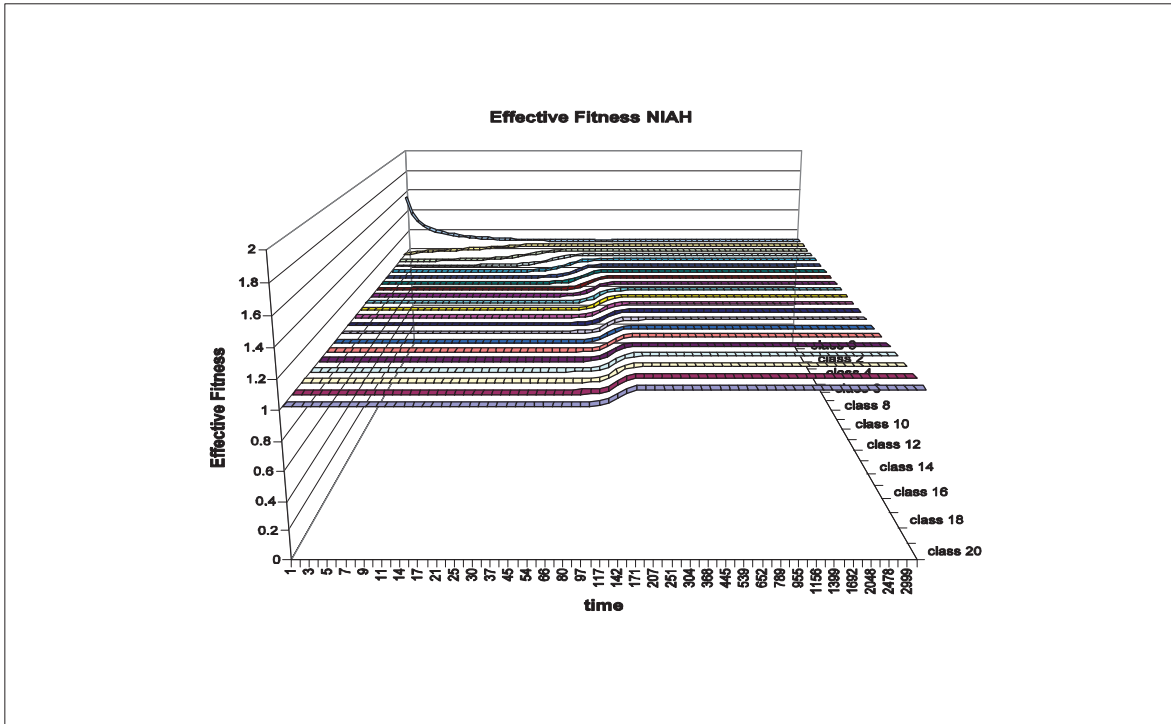
maestra ya que la aptitud efectiva de cada secuencia perteneciente a una misma clase de Hamming tiene siempre el mismo valor. Destaca que, después de la primera generación, la aptitud efectiva de la secuencia maestra (la clase 0 al fondo) disminuye mientras que la de la clase 1 aumenta. La clase 0 se estabiliza eventualmente en un valor fijo. Las demás clases presentan un aumento gradual de su aptitud efectiva hasta igualarse con la de la clase 0, sin embargo, las clases más cercanas a la clase 0 inician este aumento antes. Este aumento es provocado por el flujo de mutantes desde la secuencia maestra por tener ésta más individuos. Resulta más probable que dichos mutantes pertenezcan a la clase 1 y por eso la clase 1 se beneficia antes que las demás. Al aumentar la cantidad de individuos de la clase 1 se producen más mutantes a la clase 2 y así sucesivamente hasta la clase 20. Finalmente el paisaje de aptitud efectiva termina siendo plano lo que indica que ninguna clase tiene ventaja sobre la otra.



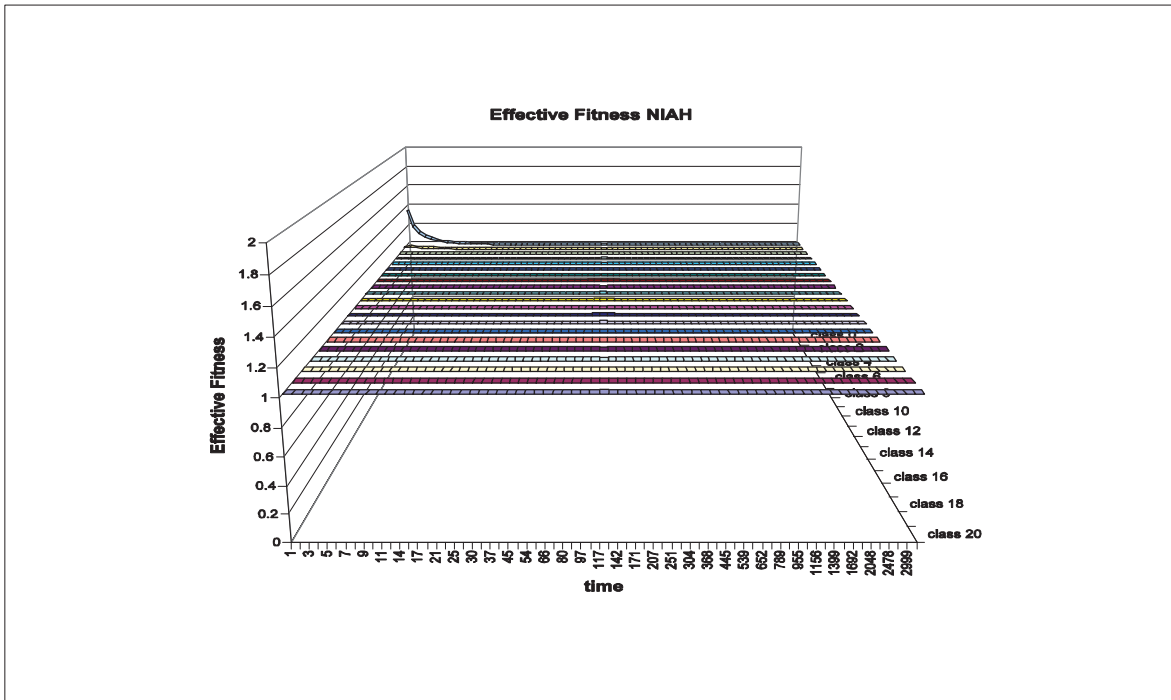
**Figura 1-8** Aptitud efectiva como función del tiempo a partir de una población inicial infinita uniforme en el paisaje de aguja en un pajar con una tasa de mutación de 0.01 que es mucho menor a la tasa crítica 0.034.

La **Figura 1-9** es análoga a la Figura 1-8 pero para una tasa de mutación mayor aunque todavía menor a la tasa crítica. Se observa una disminución importante en el nivel al cual llega la aptitud efectiva de la secuencia maestra. Esto significa que la ventaja efectiva de la secuencia maestra sobre las demás secuencias es menor.

En la **Figura 1-10** se presenta un caso donde la tasa de mutación es mayor a la tasa crítica. Aquí la ventaja de la secuencia maestra está prácticamente anulada. Esto se refleja en el hecho de que el paisaje de aptitud efectiva es prácticamente plano desde el inicio lo que significa que la evolución pierde efectivamente la señal generada por la aguja.



**Figura 1-9** Aptitud efectiva como función del tiempo a partir de una población inicial infinita uniforme en el paisaje de aguja en un pajar con una tasa de mutación de 0.03 que es un poco menor a la tasa crítica 0.034.



**Figura 1-10** Aptitud efectiva como función del tiempo a partir de una población inicial infinita uniforme en el paisaje de aguja en un pajar con una tasa de mutación de 0.04 que es mayor a la tasa crítica 0.034.



1.3.2.3 Tasa crítica de mutación  $p^*$  como función de la longitud de los genomas  $N$

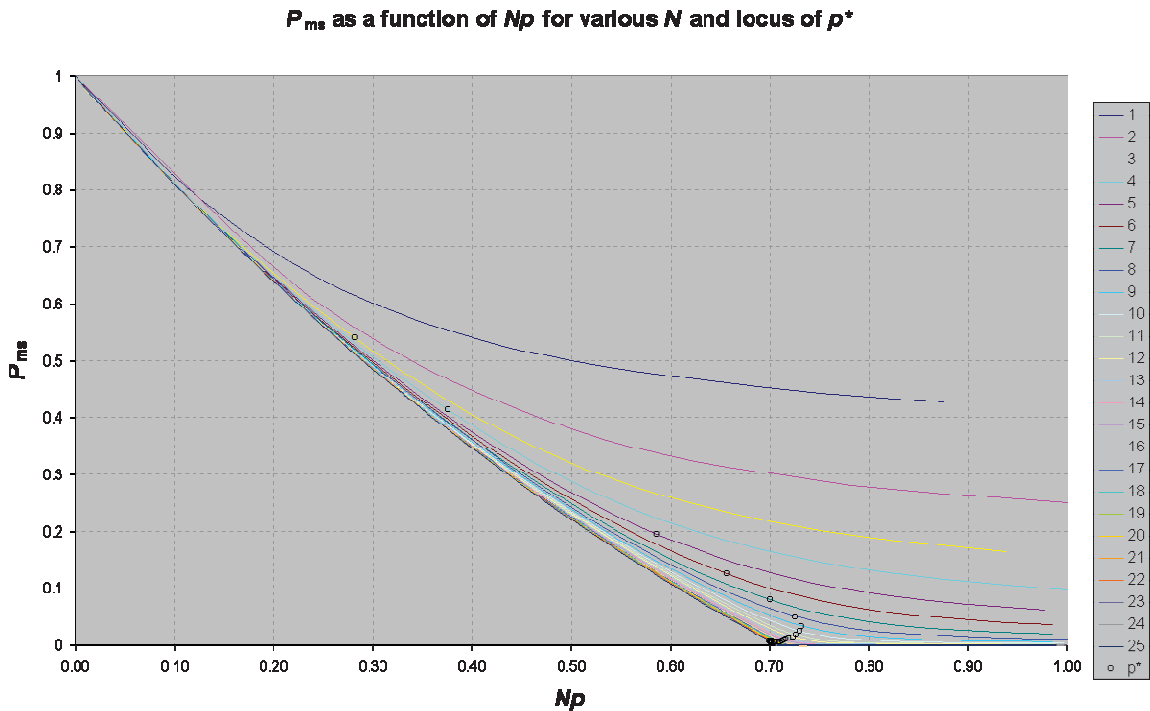
Ahora se verá cual es la relación entre la  $p^*$  y  $N$  para saber si el fenómeno de la tasa crítica de mutación está presente para toda  $N$  y cómo cambia con ésta.

El valor teórico de la tasa de mutación crítica  $p^*$  en un paisaje de aptitud de “aguja en un pajar” está dado aproximadamente por

$$p^* \approx \frac{\ln(G)}{N}, \quad G \approx \frac{f_{sm}}{f_0}$$

donde  $f_{sm}$  es la aptitud de la secuencia maestra (la aguja) y  $f_0$  es la aptitud de las demás secuencias.

Si esta aproximación es buena, entonces el producto  $Np^*$  debe ser cercano a  $\ln(G)$ . Para ver esto se integró la ecuación (1.1) para varios valores de  $N$ . En la **Figura 1-11** se muestra la gráfica de la concentración de la secuencia maestra después de 2050 generaciones (prácticamente en estado estacionario) como función del producto  $Np$  para varios valores de  $N$  (1 a 25). En estos casos  $G = 2$ .



**Figura 1-11** La concentración de la secuencia maestra como función de  $Np$  para varios valores de  $N$  y el lugar (marcado con círculos negros) donde está  $Np^*$ .

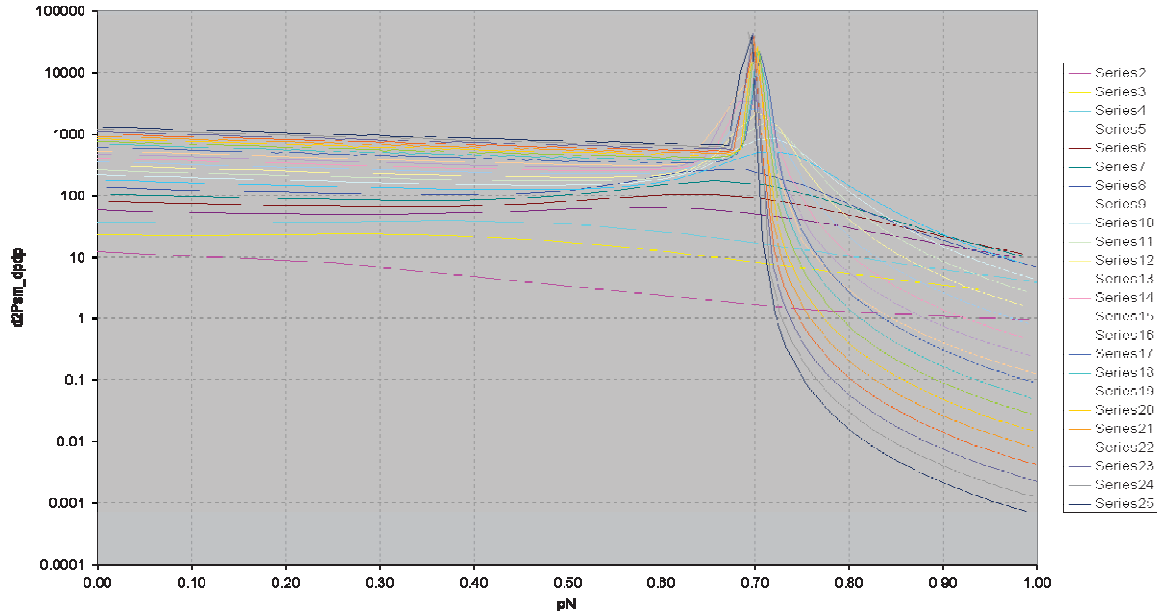


Figura 1-12 Segundas derivadas de las curvas de la Figura 1-11.

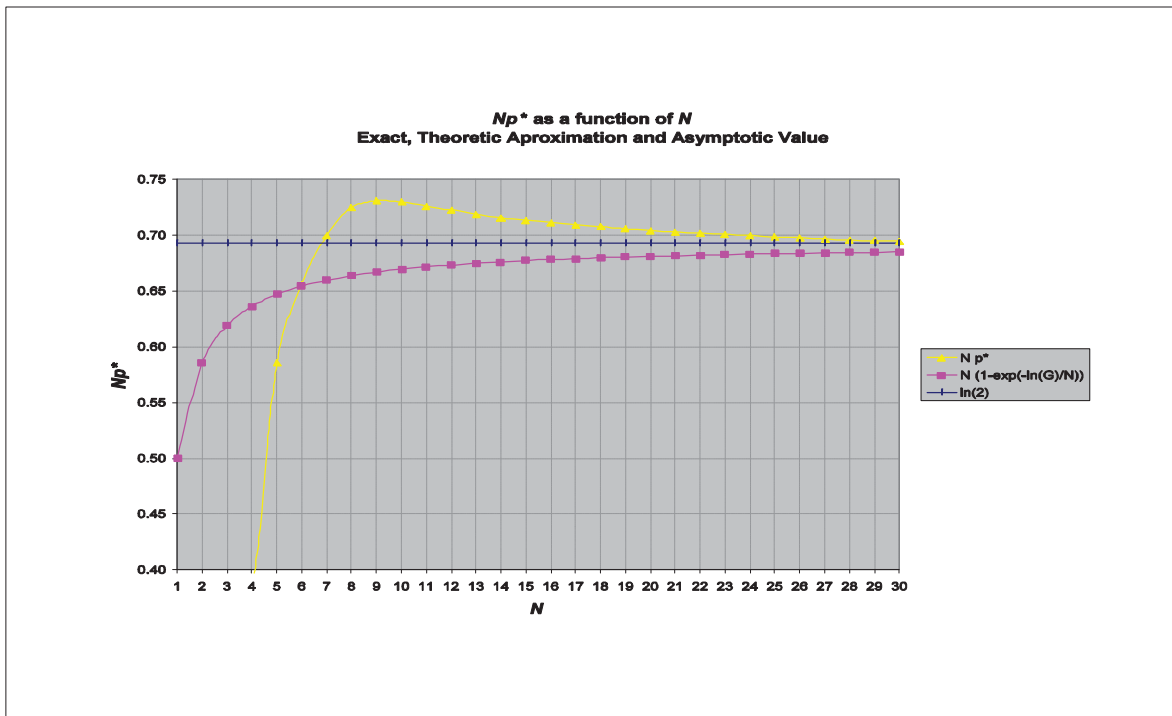


Figura 1-13  $Np^*$  como función de  $N$ , el valor de la aproximación teórica y el valor asintótico.

Puede verse en la **Figura 1-11** que a partir de aproximadamente  $Np = 0.7$  hay un cambio abrupto en el comportamiento de las curvas (para valores de  $N$  no muy bajos). Definimos a este valor como  $Np^*$ .

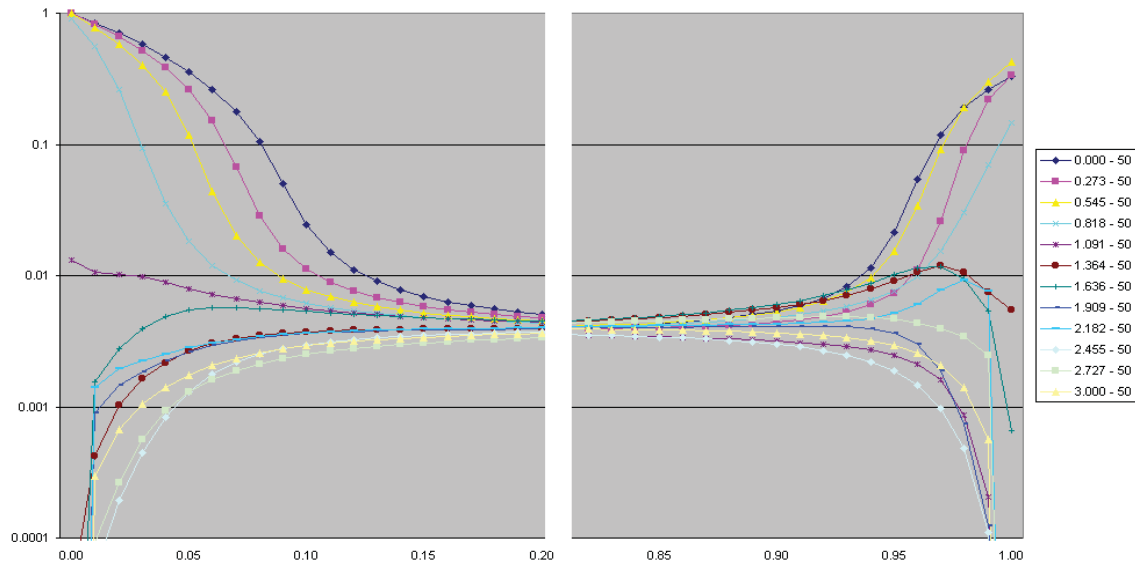
Una manera de medir con precisión en donde se da este cambio es con la segunda derivada de las curvas de la **Figura 1-11**. En la **Figura 1-12** se muestran estas segundas derivadas. El cambio brusco corresponde al punto máximo de las segundas derivadas. Se verifica que  $Np^*$  está muy cerca de 0.7.

En la **Figura 1-13** se muestra la gráfica del producto  $Np^*$  como función de  $N$  tomando la información de la **Figura 1-12**. También se muestra el valor  $\ln(2)$  para comparar los resultados con este valor que es la aproximación teórica de  $Np^*$ . Se puede ver que al aumentar  $N$  la aproximación teórica es cada vez mejor.

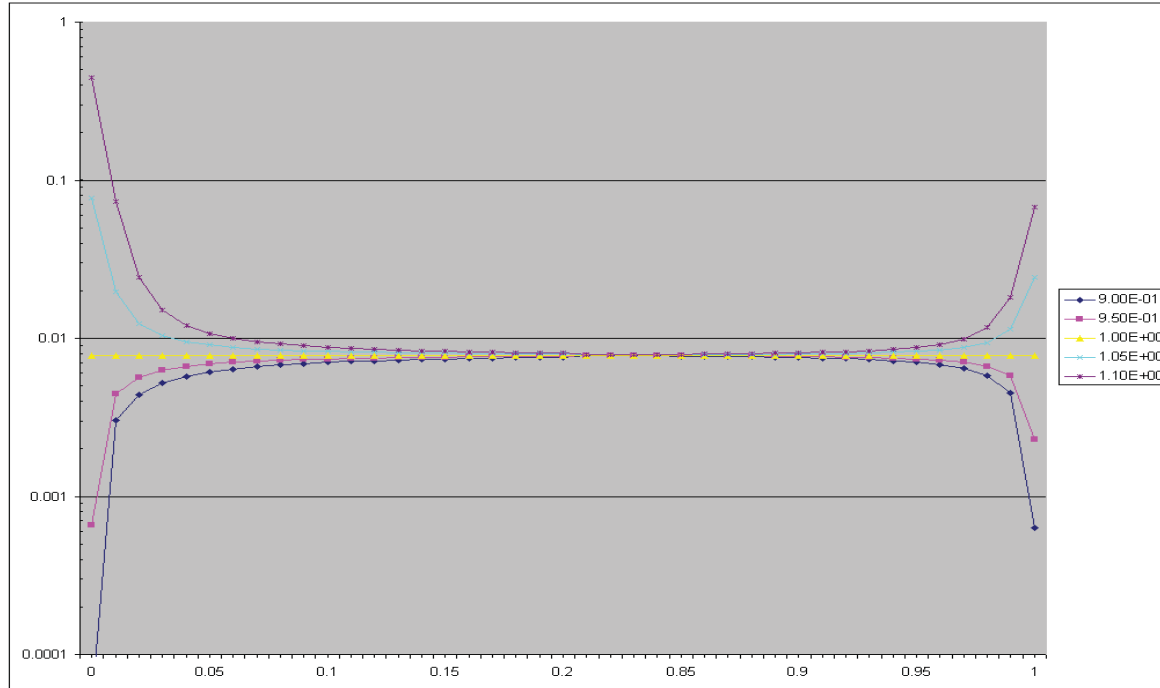
#### 1.3.2.4 *Aguja en un pajar con ruido. Universalidad del paisaje de aguja en un pajar*

Un paisaje de aptitud de aguja en un pajar tiene un pico alto en la secuencia maestra y el resto del paisaje es plano. Al introducir ruido significa que a cada secuencia, excepto a la secuencia maestra, se le agregará al azar una perturbación fija en su aptitud dentro de un rango predeterminado. Esto implica que el paisaje ya no tiene una parte plana sino que es completamente **rugoso**.

Uno se pregunta si estas perturbaciones modificarán el comportamiento general del sistema. Los resultados que a continuación se presentan demuestran, siempre y cuando el nivel de ruido no sea demasiado alto, que este comportamiento es universal, es decir, que es prácticamente el mismo.



**Figura 1-14** Detalle de  $P_{sm}$  cerca de  $p = 0$  (izquierda) y cerca de  $p = 1$  (derecha) como función de  $p$  para distintos valores de ruido máximo después de 50 generaciones. La aptitud de la secuencia maestra es siempre 2 y la de las no maestras es  $1+ruido$ . La leyenda indica el valor máximo de ruido introducido y la generación que se graficó.



**Figura 1-15** Detalle de  $P_{sm}$  cerca de  $p = 0$  (izquierda) y  $p = 1$  (derecha) para valores de  $G$  menores y mayores a uno en un paisaje de aguja en un pajar sin ruido.

En la **Figura 1-14** se muestran los resultados obtenidos con distintos niveles de ruido máximo. Para cada nivel distinto de ruido se calculó un paisaje de aptitud diferente al azar.

El comportamiento es bastante regular cuando el nivel de ruido se mantiene en valores que hacen que todas las secuencias no maestras tengan aptitudes menores que la aptitud de la secuencia maestra: mientras mayor sea el nivel de ruido, la curva para  $P_{sm}$  se traza un poco más abajo pero con la misma forma básica. Nótese también que la tasa crítica de mutación disminuye al aumentar el nivel de ruido. En cambio, cuando el nivel de ruido hace posible que algunas secuencias no maestras puedan tener mayor aptitud que la secuencia maestra, esta regularidad se pierde ya que la secuencia dominante es diferente en cada caso. En estos casos hay una tendencia de  $P_{sm}$  hacia cero cuando  $p$  tiende a cero o a uno. Esta tendencia es más pronunciada conforme aumenta el nivel de ruido máximo.

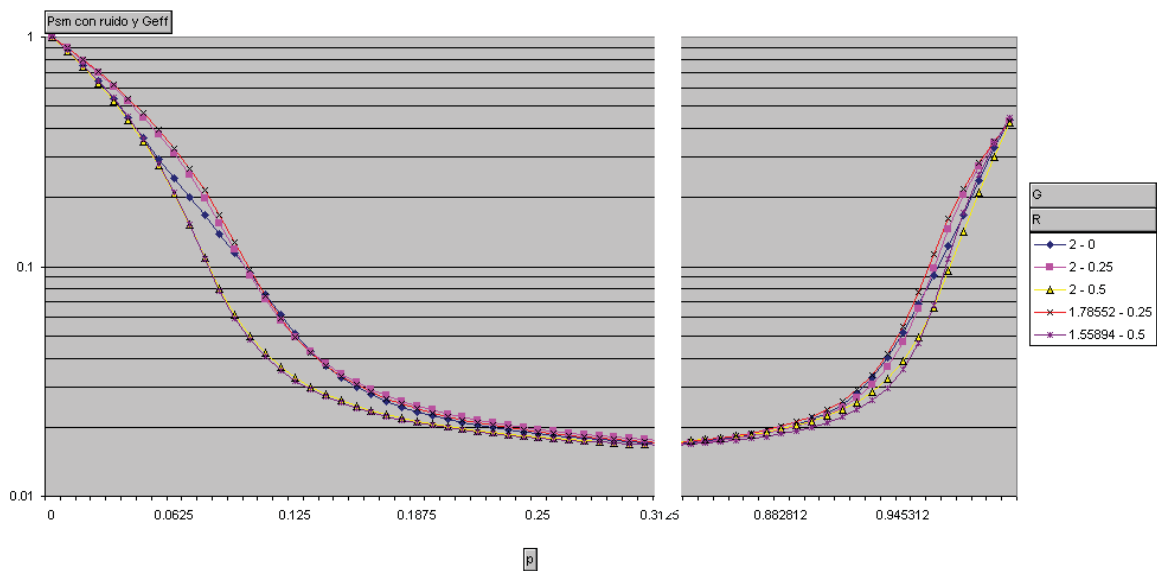
Comparando estos resultados con los de la **Figura 1-15** donde se usan diferentes valores de  $G$  desde menores a uno hasta mayores a uno sin ruido, se observa una cierta similitud en el efecto que tiene un aumento en el ruido con el efecto de un valor menor en  $G$ . Esto podría explicarse definiendo el concepto de *Ganancia equivalente promedio* como:

$$G_{eq} ? \frac{f_{sm}}{\bar{f}_{nsm}}$$

donde  $f_{sm}$  es la aptitud de la secuencia maestra y  $\bar{f}_{nsm}$  es el promedio de las aptitudes de las secuencias no maestras. En la **Figura 1-16** se muestran algunas comparaciones para diferentes

niveles de ruido contra el uso de la correspondiente ganancia efectiva sin ruido. Nótese la similitud entre la segunda y cuarta curvas y también entre la tercera y la quinta.

Cuando el paisaje de aptitud con ruido se substituye por uno con la ganancia equivalente promedio correspondiente sin ruido, se obtienen curvas muy parecidas si el nivel de ruido es relativamente pequeño. Esto indica que el comportamiento observado no solamente corresponde a un paisaje de aptitud “de juguete” sino a toda una clase de paisajes similares a éste lo cual es importante ya que es la justificación del uso de paisajes de aptitud “simples”, es decir, que el conocimiento adquirido a partir de ellos es aplicable a otros paisajes más complejos pero que tengan características similares a los ya estudiados.



**Figura 1-16** Comparación entre paisajes de aptitud con ruido y paisajes con Ganancia equivalente promedio sin ruido, con  $p$  cerca de 0 (izquierda) y de 1 (derecha). Las primeras tres leyendas indican la ganancia real  $G$  y el nivel de ruido usado. Las otras dos leyendas indican la ganancia efectiva usada y el nivel de ruido del cual se calculó esa ganancia. La primera curva es la que no tiene ruido.

## 1.4 La Cuasiespecie y agujas degeneradas

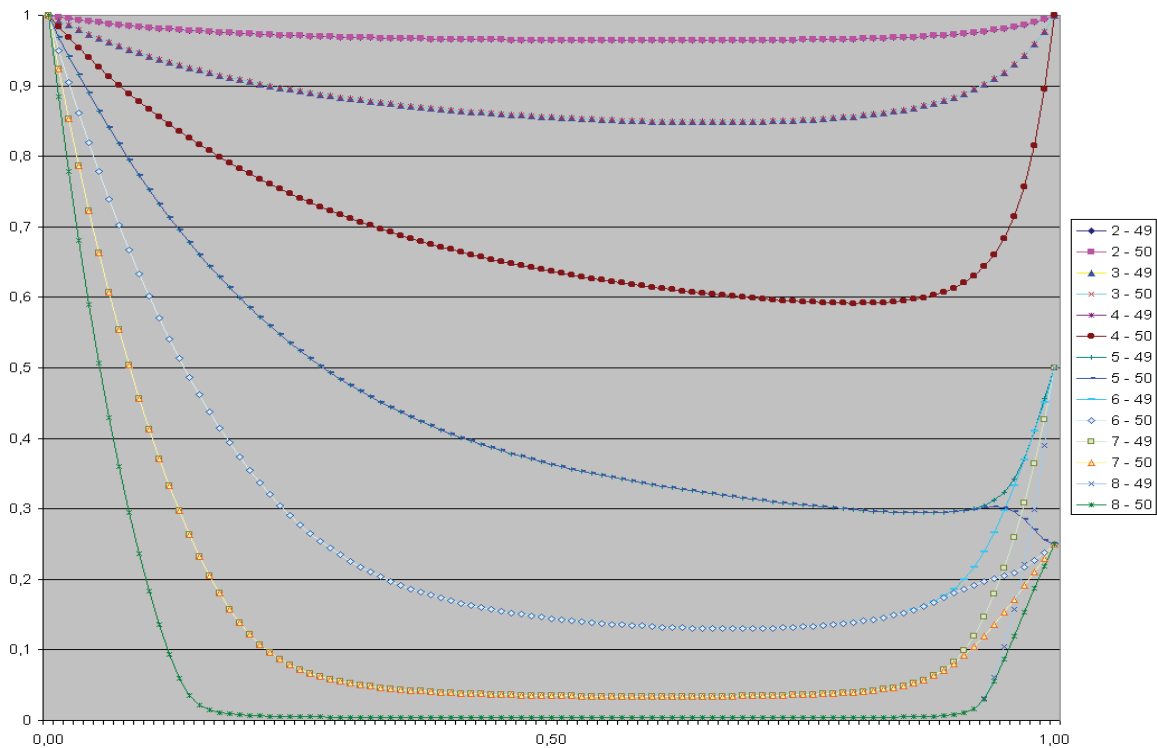
Ya se ha visto hasta aquí cómo interactúan entre sí la mutación y la selección en los paisajes plano y con una secuencia maestra solamente. El siguiente paso es analizar el sistema con un paisaje de aptitud en el que no solo hay una sino varias secuencias maestras que tienen una aptitud superior a las demás. Todas estas secuencias maestras tienen la misma aptitud así que el paisaje presenta degeneración tanto en las secuencias no óptimas como en las óptimas.

Se estudian a continuación diferentes maneras de introducir degeneración en las secuencias

maestras. La más simple es que todas las secuencias maestras sean contiguas (a una distancia de Hamming 1) en el espacio de los genotipos. Otra posibilidad es que se distribuyan en el espacio de los genotipos un número de secuencias maestras y así estudiar la dependencia que hay en el comportamiento del sistema respecto a la geometría de tal distribución.

### 1.4.1 Degeneración con secuencias maestras contiguas

En este caso se requiere que las secuencias maestras estén todas agrupadas alrededor de una secuencia central y que todas tengan al menos una secuencia maestra como vecina a una distancia de Hamming 1 de ellas. Para hacer esto en el modelo se consideró que aquellas secuencias que contengan por lo menos  $L$  genes iguales a 1 tendrían un fenotipo de secuencia maestra y los demás de secuencia no maestra. Entonces si, por ejemplo, se considera  $N = 7$  y  $L = 7$ , se tiene solamente una secuencia maestra, si  $N = 7$  y  $L = 6$ , se tienen 8 secuencias maestras, etcétera.



**Figura 1-17**  $P_{sm}$  como función de  $p$  para diferentes valores de  $L$  en las generaciones 49 y 50 con  $N = 8$  y  $G = 3$ .

Se muestra en la **Figura 1-17** la gráfica de  $P_{sm}(p)$  para  $G = 3$ ,  $N = 8$  y  $L = 2, 3, 4, 5, 6, 7$  y  $8$  en las generaciones 49 y 50 iniciando con una población uniforme. En estos casos  $P_{sm}$  es la fracción de la población que ocupan todas las secuencias maestras.

Como se vio en la sección 1.2.2 cuando  $p = 1/2$  se tiene una población uniforme. En cada caso de la **Figura 1-17** con  $p = 1/2$ , se tiene que  $P_{sm}$  es exactamente el número de secuencias

maestras entre el número de posibles secuencias distintas.

Todas las curvas coinciden en  $p = 0$  ya que, en este caso, solamente las secuencias maestras proliferan en la población y las demás se extinguen. Cuando  $p$  tiende a 1, se observan claramente dos comportamientos distintos. Para  $L = 1, 2, 3, 4$   $P_{sm}$  tiende a uno pero para  $L = 5, 6, 7, 8$  se observa que en la generación 49 las curvas van hacia  $1/2$  y para la generación 50 las curvas van hacia  $1/4$ .

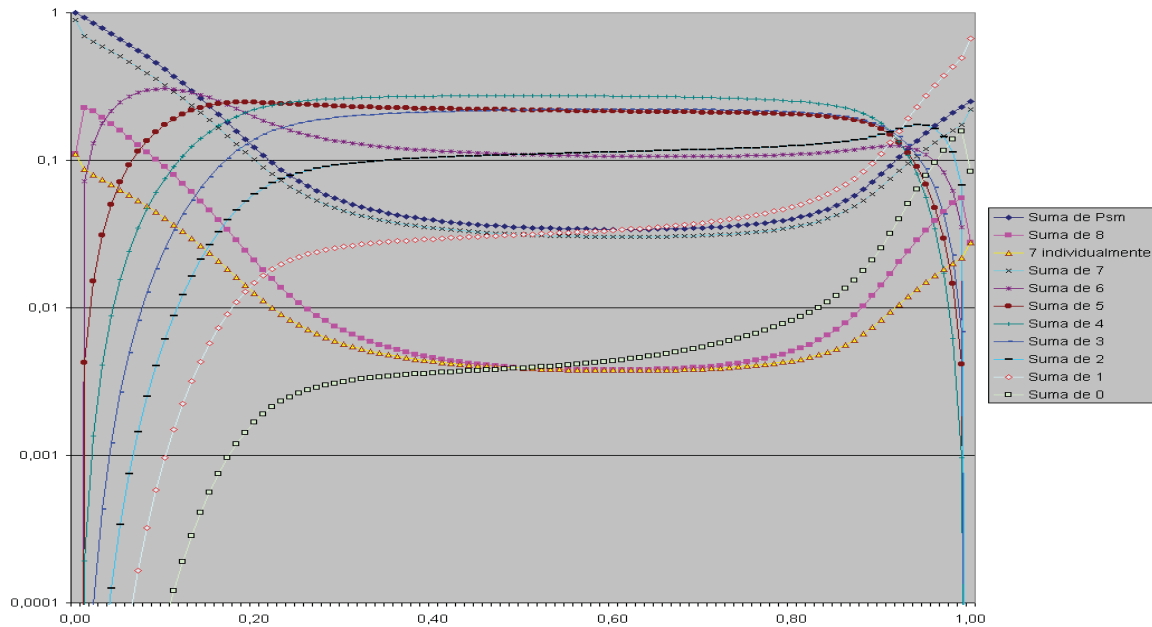
El primer caso se explica porque, al requerirse menos de la mitad o la mitad de los 8 genes (bits) en 1, los demás genes (que son 4 o más) pueden tener cualquier valor incluyendo a puros ceros. Así, como  $p$  está cerca de 1, hay una alta probabilidad de que una secuencia se convierta en su opuesta y, por lo tanto, de que se lleve a las secuencias maestras con 4 o más ceros a convertirse en otra secuencia maestra. De esta manera, las secuencias maestras con 4 ceros y 4 unos dominan a toda la población después de algunas generaciones, ya que, en todas las generaciones la selección las favorece.

Por el otro lado, cuando  $L > 4$ , los genes (bits) restantes son menos que los genes requeridos en uno, así que la mutación tiene una muy baja probabilidad de convertir una secuencia maestra a otra secuencia maestra. Sin embargo, como  $p$  es cercana a 1, hay una alta probabilidad de que las secuencias maestras muten en una secuencia no maestra que tiene una alta probabilidad de mutar en una secuencia maestra (ella misma). Por lo tanto se presenta una *oscilación* entre secuencias maestras y sus opuestas (que no son maestras) la cual tiende a ser cada vez de menor amplitud al paso de las generaciones. Mientras más cerca esté  $p$  de 1, dicha reducción en la amplitud de la oscilación será más lenta y, justo cuando  $p$  es 1, la oscilación no se reduce. Los pares de secuencias opuestas formados por una secuencia maestra y una no maestra dominan a los pares de secuencias opuestas formados por dos secuencias no maestras y, como todas las secuencias maestras que quedan en la población y sus opuestas tienen las mismas aptitudes, se comportan como si se tratara de una sola secuencia maestra y su opuesta, tal y como sucede en el caso de un gen de la sección 1.2.5. Esto último explica el porqué las curvas para  $L = 5, 6, 7, 8$  coinciden en  $p = 1$ . Los valores en los que coinciden ( $1/2$  y  $1/4$  en este caso) dependen del valor de  $G$  como en el caso de un gen.

En la **Figura 1-18** se muestran las concentraciones de cada clase de *Hamming* cuando  $N = 8$ ,  $L = 7$  y  $G = 3$  después de 49 generaciones. Se observa que, cuando  $p = 0$ , las clases de Hamming 7 y 8 dominan toda la población y cuando  $p = 1$  dominan también las clases 7 y 8 pero junto con sus opuestas que son las clases 1 y 0 respectivamente. Cuando  $p > 0$  y  $p < 1$ , la concentración de la secuencia con 8 genes en uno es mayor que las concentraciones de cada una de las 8 secuencias con 7 genes en uno. Esto se debe a que las secuencias que “rodean” a la de 8 (las que están a una distancia de Hamming 1) son secuencias muy aptas (las de 7) mientras que las de 7 están rodeadas en promedio por secuencias menos aptas, entre ellas está la de 8 pero las demás son secuencias no maestras. Se dice entonces que en este caso ha habido un **rompimiento inducido de la simetría** en el paisaje con el cual se da una distinción entre las secuencias maestras a pesar de que tienen el mismo fenotipo o aptitud.

Nótese en la **Figura 1-18** que la clase 8 (una sola secuencia) y la clase 7 individualmente coinciden en  $p = 0$ . Esto se debe a que con  $p = 0$  todas las secuencias no maestras tienden a desaparecer quedando solamente secuencias maestras formándose un paisaje de aptitud plano

entre ellas. Teniendo  $p = 0$  y un paisaje plano, se mantiene en toda generación las mismas concentraciones de cada secuencia que se tenían inicialmente. Como se usó una población inicial uniforme, deben entonces coincidir dichas curvas. Sin embargo, estas curvas son muy distintas cuando  $p$  es apenas un poco mayor que cero. Esto se debe a que el efecto de la población inicial ya no es infinito y, de hecho, tiende a desaparecer, aunque sea después de muchas generaciones.



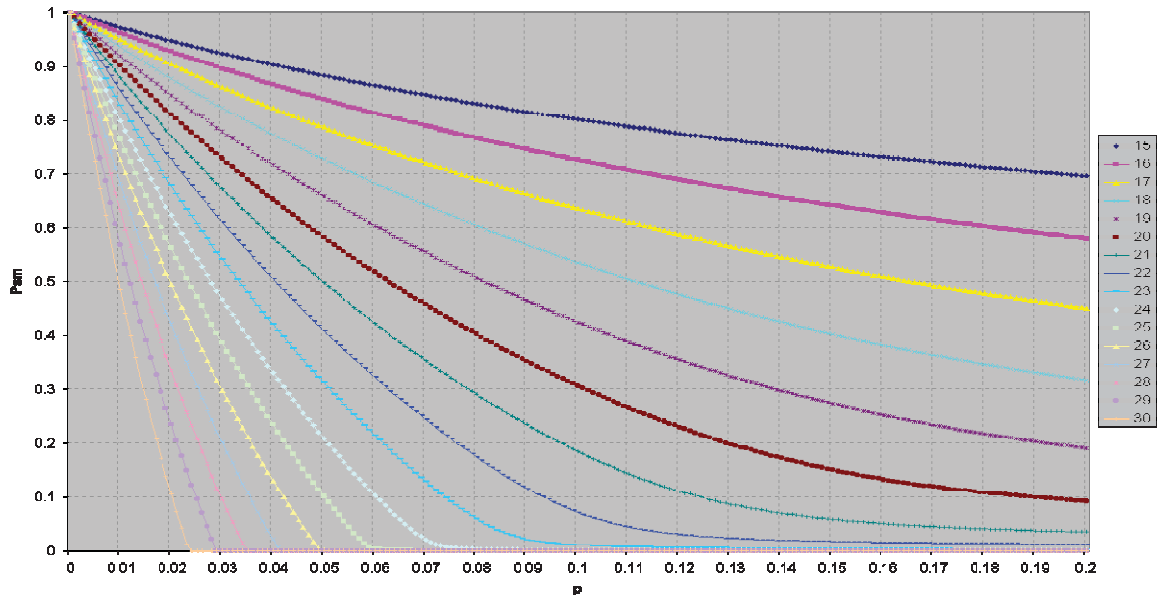
**Figura 1-18** Gráfica de  $P_{sm}$  cuando  $N = 8$ ,  $L = 7$ ,  $G = 3$  y de cada clase de Hamming como función de  $p$  en la generación 49. También se muestra la concentración de cada una de las secuencias con 7 genes en uno individualmente para que pueda compararse con la concentración de la secuencia con 8 genes en uno.

#### 1.4.1.1 Tasa de mutación crítica en función del grado de degeneración de la aguja

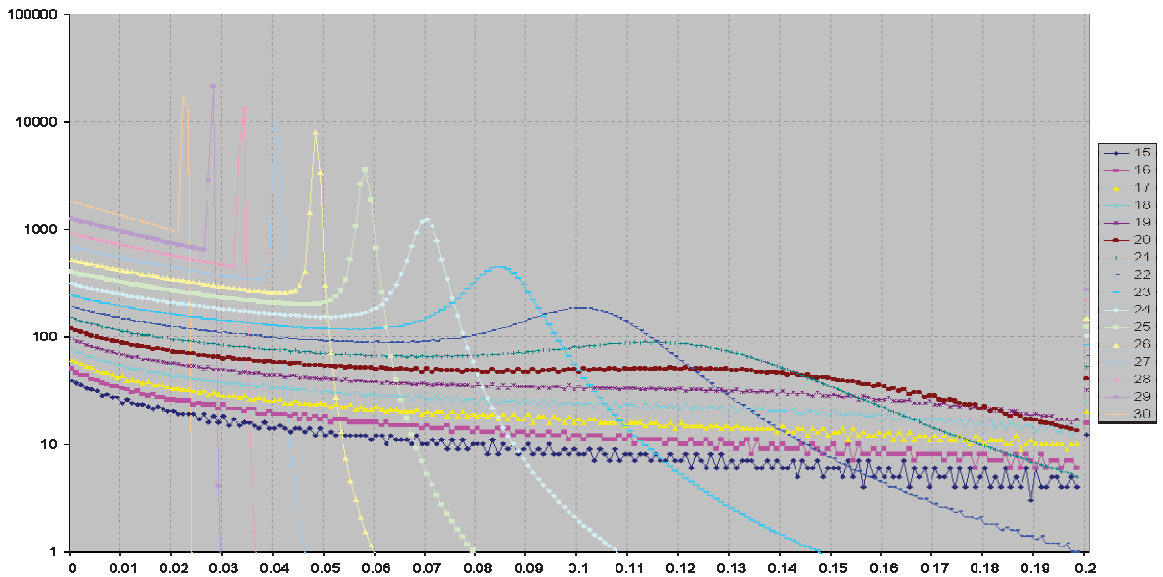
Cuando se hace variar el parámetro  $L$  (número de genes iguales a la secuencia maestra requeridos para que la aptitud de una secuencia sea igual a la de la secuencia maestra) se tiene una “señal” en el paisaje más o menos ancha (en cuanto al número de secuencias maestras), es decir, que se forma una red neutral de secuencias degeneradas con la misma ventaja de aptitud sobre el resto. ¿Cómo varía la tasa de mutación crítica  $p^*$  en función de este grado de degeneración en la aguja?

En la **Figura 1-19** puede verse que para valores grandes de  $L$  el valor de  $p^*$  es menor. Sin embargo para valores bajos de  $L$  cada vez es más difícil determinar cual es el valor de  $p^*$ . Esto es porque las curvas son cada vez más suaves y el cambio abrupto en el comportamiento de éstas ya no es tan abrupto. Incluso llega a desaparecer implicando la no existencia de una  $p^*$ .





**Figura 1-19** Concentración de las secuencias maestras como función de  $p$  (valores bajos) para varios valores de  $L$  con  $N = 30$ .



**Figura 1-20** Segundas derivadas de las curvas de la Figura 1-19.

Para medir con mayor precisión en donde se da la  $p^*$  se puede usar la segunda derivada de las curvas de la **Figura 1-19**. La  $p^*$  es el punto donde la segunda derivada tiene un pico máximo local. Estas segundas derivadas se ven en la **Figura 1-20**.

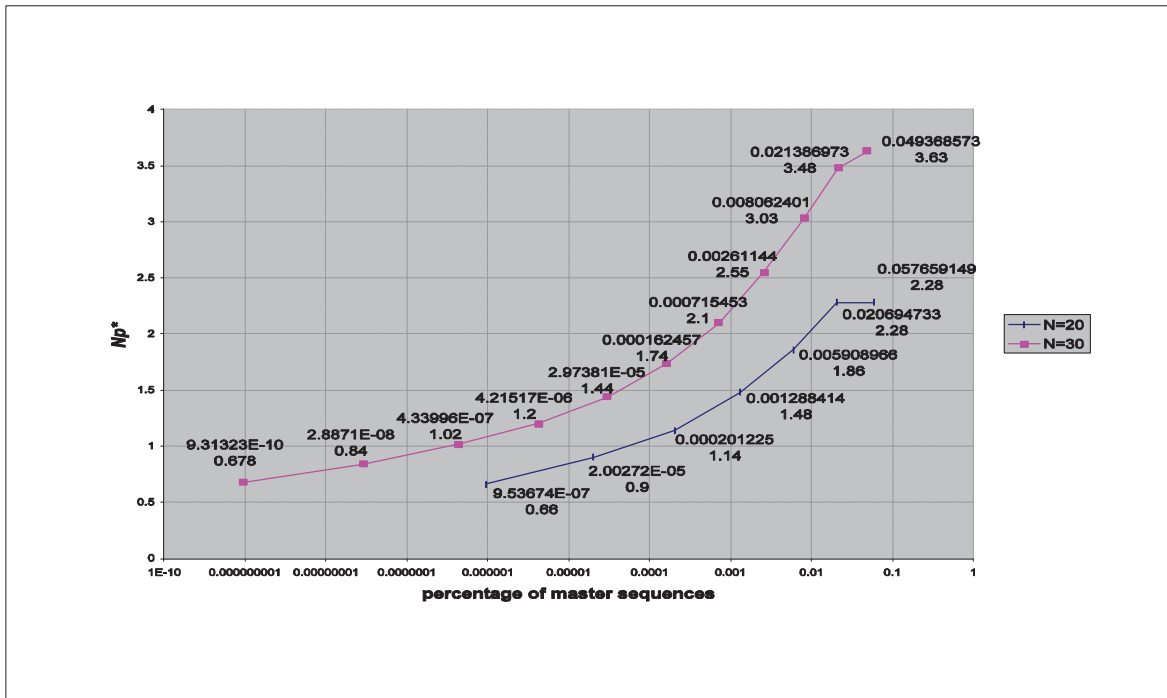


Figura 1-21  $p^*$  como función del número de secuencias maestras.

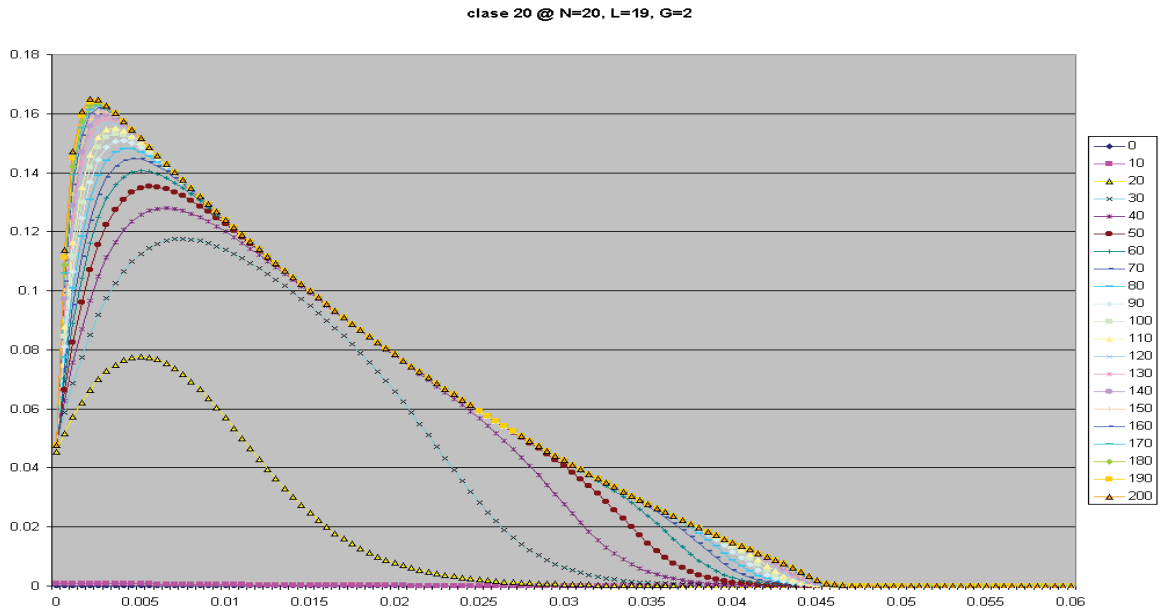
De las curvas de la **Figura 1-20** se obtiene la información para hacer la gráfica de la **Figura 1-21** en la que se muestra la  $p^*$  como función de la fracción del espacio de búsqueda que ocupan las secuencias maestras para  $N = 20$  y  $N = 30$ . Los puntos de las curvas fueron tomados dando valores distintos a  $L$  y calculando después la fracción correspondiente a cada valor.

De la **Figura 1-21** se desprende que hay una relación universal entre el grado de degeneración de la aguja y la tasa crítica de mutación. Nótese como las curvas inician desde el valor teórico para agujas no degeneradas ( $\ln(G)$  con  $G = 2$ ) y terminan en donde la aguja ocupa aproximadamente el 5% del espacio de búsqueda. Más allá de este 5% la tasa crítica desaparece por completo, es decir, que ya no se observa un pico en la segunda derivada de las concentraciones respecto a  $p$ .

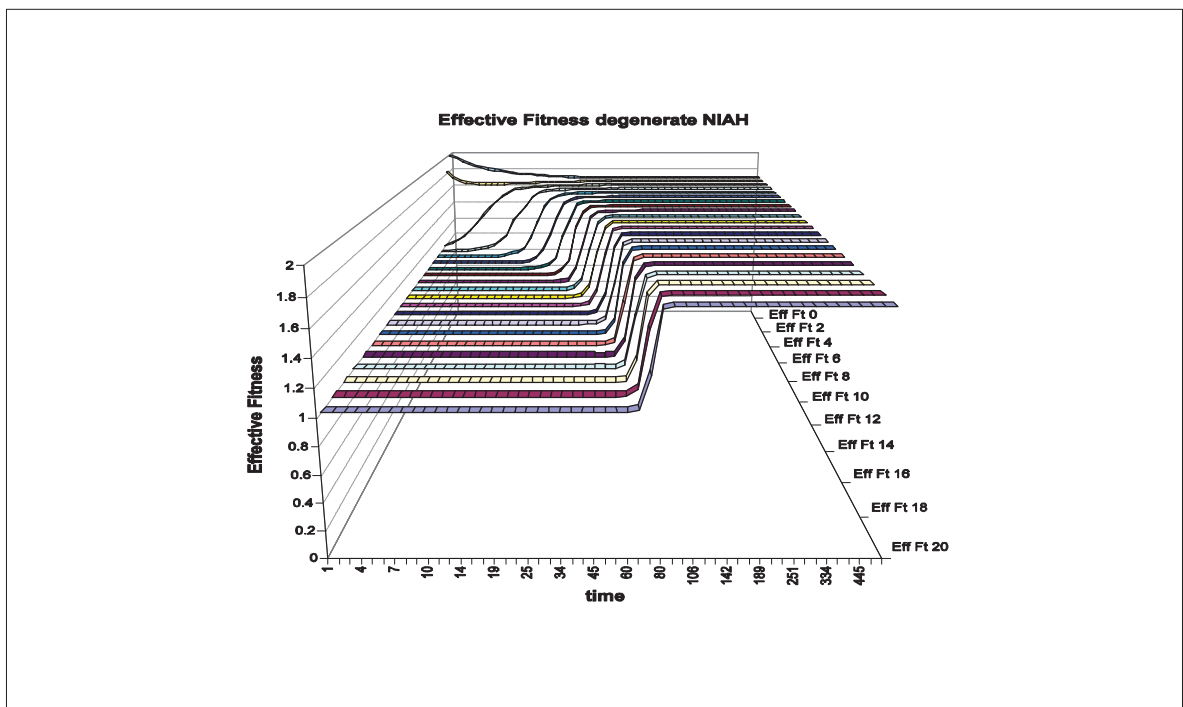
El que  $p^*$  aumente con el grado de degeneración de la aguja implica que ésta, la aguja, será más robusta ante el efecto de la mutación ya que se requieren tasas más altas para que se extinga la cuasiespecie alrededor de la aguja. Esto explica en parte porque en la naturaleza se da con tanta frecuencia la degeneración.

#### 1.4.1.2 La formación de una Cuasiespecie a través del tiempo

Hasta este punto se ha estudiado cuál es el punto de equilibrio del sistema pero, ¿Cómo se llega hasta allí? En esta sección se ve precisamente esto y además las implicaciones que pueden derivarse del comportamiento observado para sistemas con población finita.



**Figura 1-22** En esta gráfica se muestra la concentración de secuencias de la clase 20 como función de  $p$  cerca de  $p = 0$  para distintos valores de  $t$  (generaciones) de una población de individuos con  $N = 20$  y  $L = 19$ .

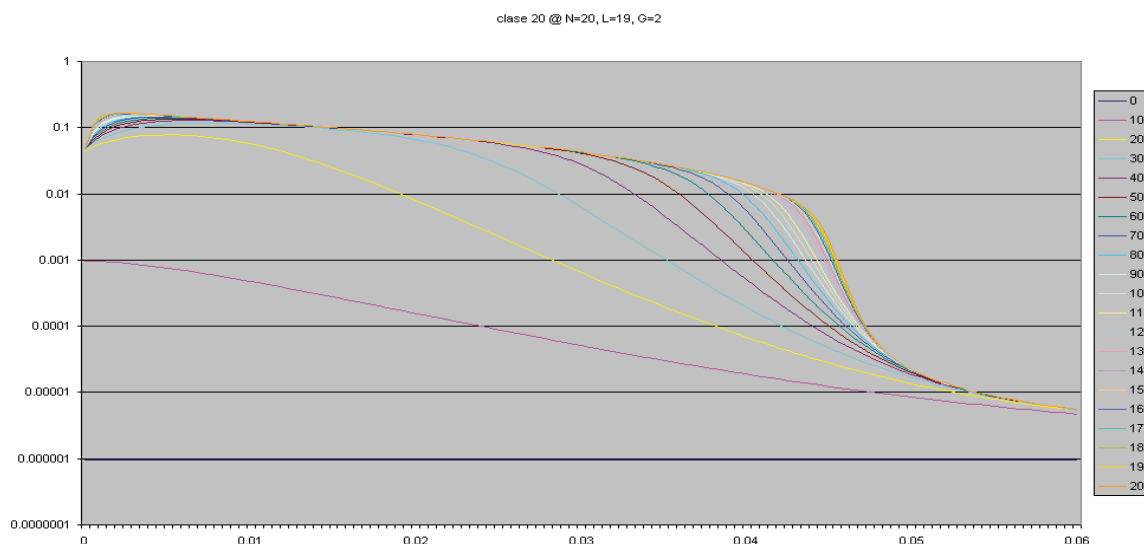


**Figura 1-23** Aptitud efectiva de cada clase de Hamming en el paisaje de aguja degenerada con una  $p$  baja. Las clases 20 y 19 son clases maestras con  $G = 2$ .  $N = 20$ .

En la **Figura 1-22** se muestra el progreso a través del tiempo de la clase 20 (una sola

secuencia), a partir de una población inicial uniforme cuando  $N = 20$ ,  $L = 19$  y  $G = 2$  desde  $p = 0$  hasta un poco más allá de la  $p$  crítica. Cada curva corresponde con un valor de tiempo (generaciones) en función de la tasa de mutación  $p$ . En  $t = 0$  la curva es horizontal y al pasar el tiempo las curvas son cada vez más altas.

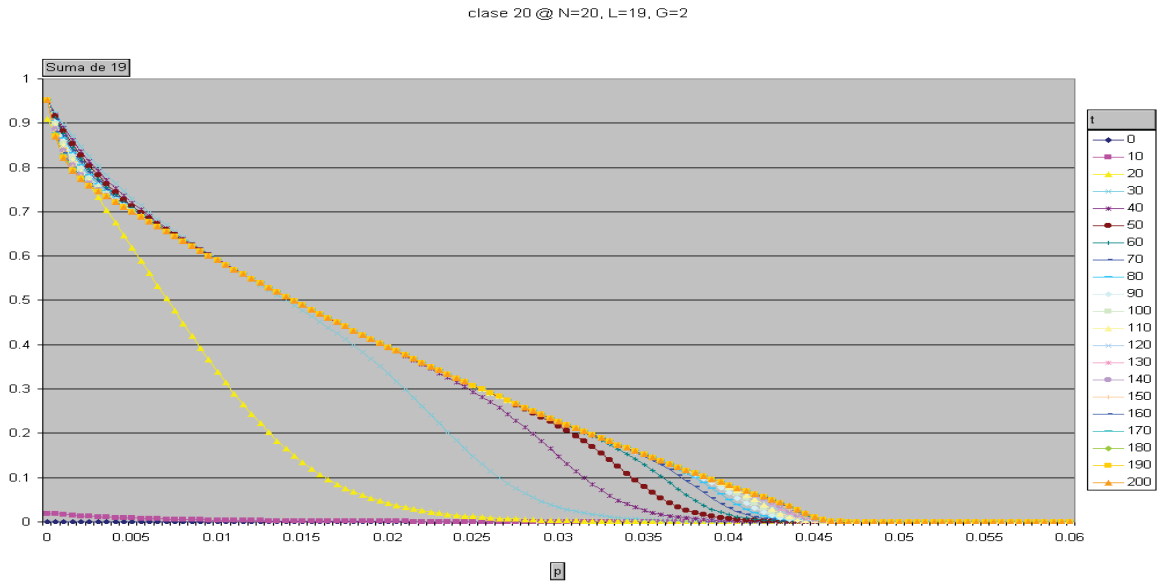
Se observa que, para algunos valores de  $p$ , la curva converge rápidamente al valor en estado estacionario. Esto es para valores de  $p$  aproximadamente entre 0.015 y 0.025. Cerca de  $p = 0$ , la concentración aumenta primero rápidamente hasta un valor un poco mayor a 0.04, que es cuando las secuencias maestras terminan de dominar a la mayoría de las secuencias no maestras, y luego sigue aumentando más lentamente. Esta segunda fase corresponde a la competencia que se da entre las clases de secuencias maestras 20 y la 19. La clase 20 tiende a dominar a la clase 19 ya que a esta última la forman secuencias que tienen como vecinas algunas secuencias no maestras lo cual reduce su aptitud efectiva. Véase en la **Figura 1-23** cómo la clase 20 (al fondo) tiene siempre mayor aptitud efectiva que la clase 19. La aptitud efectiva muestra claramente la formación de la Cuasiespecie, que se inicia en la aguja y va creciendo desde allí hacia el resto del espacio de los genotipos pasando primero por los más cercanos en términos de la distancia de Hamming.



**Figura 1-24** Esta gráfica es la misma que la de la *Figura 1-22* pero en escala logarítmica para mostrar el detalle cuando  $p$  es cercana a  $p^*$ .

La **Figura 1-24** es la misma que la **Figura 1-22** pero en escala logarítmica para mostrar con más detalle el comportamiento para  $p$  cerca de  $p^*$ . Nótese la fuerte caída cerca de  $p^*$ . Conforme  $p$  se acerca a la  $p$  crítica (desde valores menores), la convergencia se hace también más lenta debido a la fuerte competencia entre la presión selectiva y la mutación. Esto hace que la clase 20 no domine a la población (ni la 19) durante varias generaciones. Mientras las secuencias maestras no dominen, se mantendrá una población casi uniforme. Este efecto es importante ya que, en una población finita, la población de secuencias maestras puede extinguirse debido a que en las primeras generaciones se mantiene un número bajo de ellas aumentándose la probabilidad de que se extingan por mutaciones. Véase la sección 1.7.1.2 en

la página 83 acerca de búsqueda genética con población finita.



**Figura 1-25** La clase 19 cuando  $N = 20$ ,  $L = 19$  como función de  $p$  para varios valores de  $t$ .

En la **Figura 1-25** se muestra ahora el comportamiento de la clase 19. Se observa que, cuando  $p$  es cercana a cero, las curvas se alzan primero rápidamente y luego la evolución se hace lenta pero, a diferencia de lo visto en la **Figura 1-22**, aquí la mutación desfavorece a la clase 19 frente a la clase 20 y poco a poco disminuye la concentración de esta clase cediendo terreno a la clase con mayor aptitud efectiva, la clase 20. Cerca de la  $p^*$  se observa el mismo comportamiento que en la clase 20.

El comportamiento visto en estos casos cerca de la  $p^*$  no se presenta en el caso de un gen y la razón es la siguiente. De la **Figura 1-2** en la sección 1.2.5 puede concluirse que al menos con  $N = 1$  siempre se llega al punto estable más pronto conforme  $p$  se acerca a  $\frac{1}{2}$  ya que el cociente  $\frac{p}{1-p}$  se aproxima cada vez más a 0. Esto se debe a la forma que toma la curva a la que se converge. Véase en la **Figura 1-26** que en ese caso ( $N = 10$ ) la curva de convergencia es mucho más suave que, por ejemplo, la de la **Figura 1-24**. Para valores de  $N$  menores, (Véase la **Figura 1-27**) esta curva es aún más estirada llegando al punto en el que las curvas para diferentes valores de  $t$  “llegan” a tocarla siempre de derecha a izquierda, lo que significa que a mayor  $p$  se tendrá siempre menor tiempo para llegar al nivel de convergencia. Se concluye que el efecto descrito para  $N = 20$  cerca de la  $p$  crítica solo está presente para valores de  $N$  grandes y que mientras más grande sea  $N$  el efecto será más notorio.

Se ha dicho en la literatura que una buena tasa de mutación es aquella que mantiene al sistema “al borde del caos” y que esa tasa está relacionada con la tasa crítica de mutación  $p^*$ . Con lo que se ha visto aquí se pone en duda esta aseveración ya que si la  $p$  está cerca de la  $p^*$  se tiene una alta probabilidad de que la secuencia maestra quede extinta, antes de poder crear una Cuasiespecie robusta a partir de la cual se siga haciendo la búsqueda, lo cual puede resultar ser

una gran perdida si la secuencia maestra es difícil de encontrar nuevamente. Todo esto se analiza con más detalle en la sección 1.7.1.2 en la página 83.

$N=10, L=9$

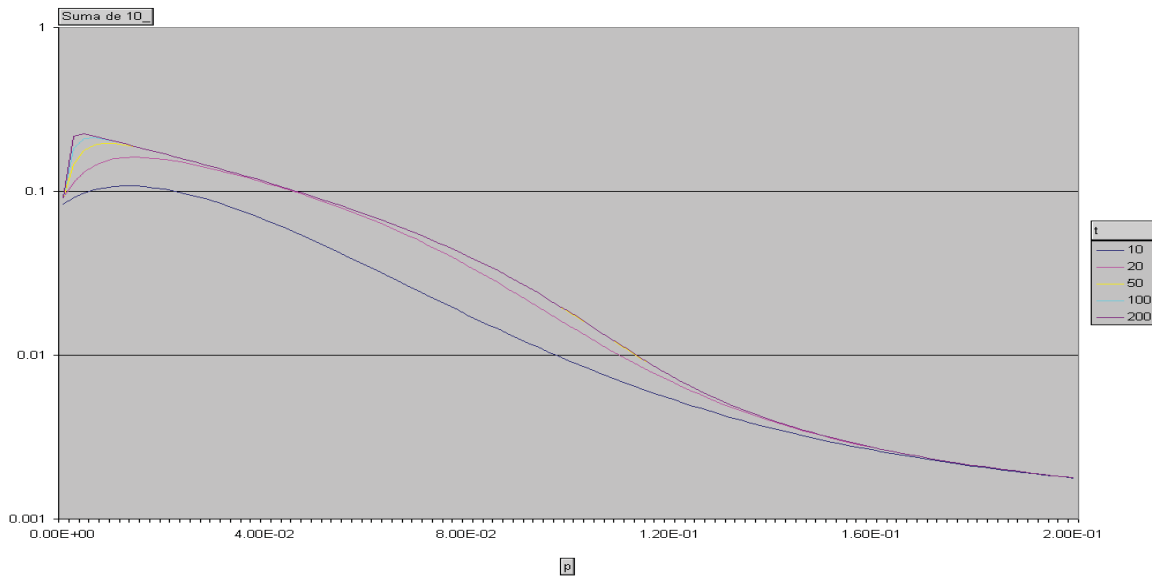


Figura 1-26 La clase 10 cuando  $N = 10, L = 9$  como función de  $p$  para varios valores de  $t$ .

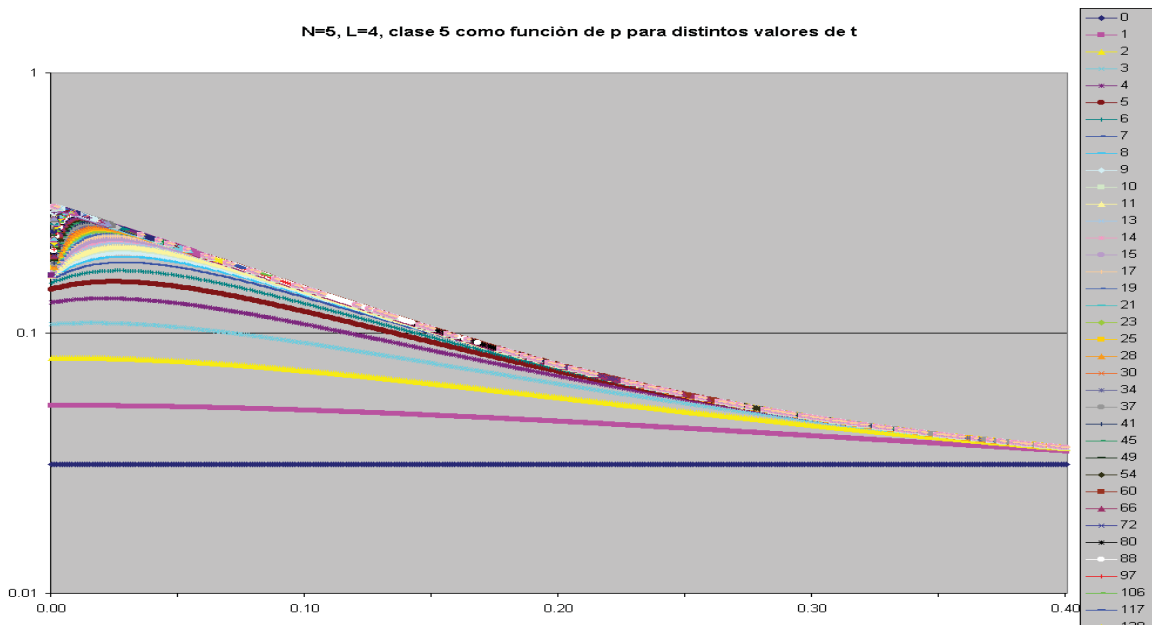
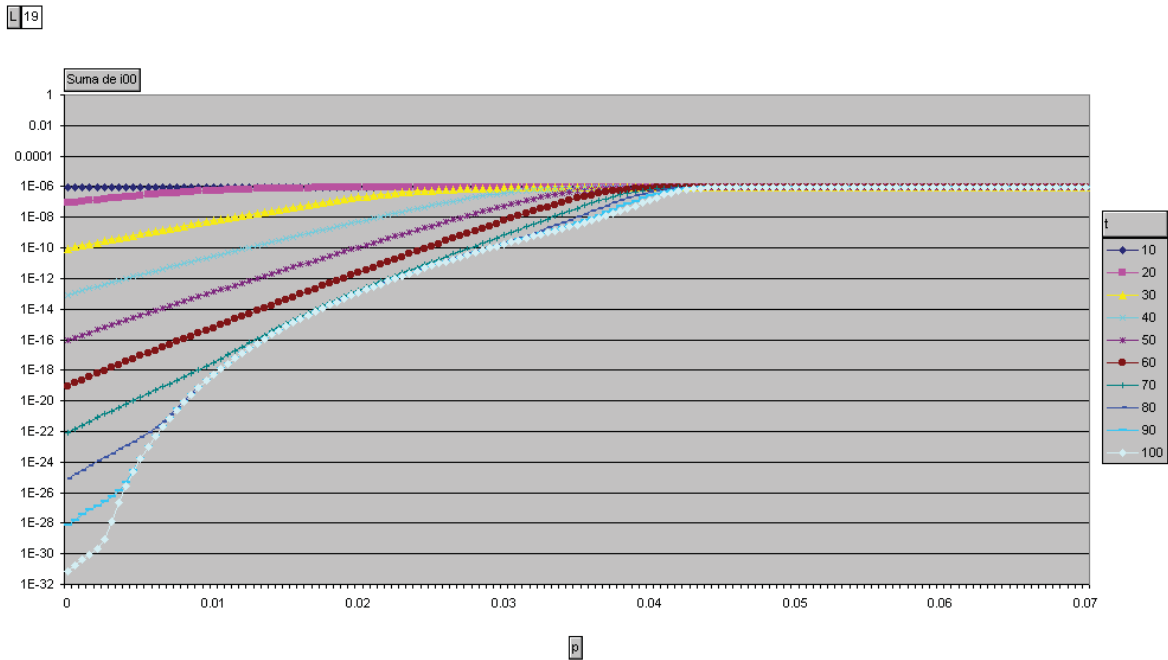


Figura 1-27 La clase 5 cuando  $N = 5, L = 4, G = 2$  como función de  $p$  para distintos valores de  $t$ .

En la **Figura 1-28** se muestra el progreso de la concentración de la clase 0 en el mismo caso de la **Figura 1-22**. Se ve cómo es que las curvas progresan de una manera simple hacia abajo hasta que llegan a la curva de convergencia donde simplemente dejan de avanzar. En ese

punto las clases maestras producen suficientes mutantes hacia la clase 0 como para mantener ese nivel de población.



**Figura 1-28** La clase 0 cuando  $N = 20$ ,  $L = 19$  como función de  $p$  para varios valores de  $t$ .

El comportamiento observado en las figuras anteriores puede explicarse de la siguiente manera. Se presentan dos fenómenos superpuestos en la formación de la Cuasiespecie. Uno tiene que ver con el dominio de las secuencias maestras sobre las no maestras y el otro con el dominio de unas secuencias maestras sobre otras secuencias también maestras.

El primero depende de qué tantas secuencias maestras quedan después de aplicar selección y mutación. Cuando hay poca mutación quedan casi todas las secuencias maestras y por lo tanto hay un dominio rápido sobre las no maestras. Al aumentar la mutación quedan cada vez menos secuencias maestras y más secuencias no maestras, reduciéndose la ventaja efectiva de las primeras sobre las segundas. Se llega a un punto en el que la ventaja es prácticamente nula, que es cuando  $p$  es la tasa crítica  $p^*$ . El nivel de  $P_{sm}$  (fracción de secuencias maestras en la población) en el punto de equilibrio (o nivel de convergencia) se va haciendo menor al aumentar  $p$  ya que en cualquier tiempo  $t$  se tendrá una cantidad cada vez mayor de mutantes que son secuencias no maestras.

El segundo fenómeno depende de cuántas secuencias maestras “rodean” a cada secuencia maestra. Las secuencias maestras que están rodeadas por más secuencias maestras dominan a las que están rodeadas por menos secuencias maestras. La velocidad a la que se da este dominio depende de  $p$ . Cuando hay poca mutación, simplemente el número de mutantes es muy pequeño y por lo tanto en cada generación aparecerán muy pocos mutantes que sean secuencias no maestras así que la ventaja que tienen las secuencias maestras rodeadas de buenas secuencias se presenta muy lentamente. Al aumentar la tasa de mutación  $p$ , el proceso

se acelera y las secuencias maestras rodeadas por secuencias maestras ejercen su dominio más rápidamente simplemente porque reciben más mutantes desde sus vecinas que las que tienen malas vecinas. Al ir dominando algunas secuencias maestras a otras, habrá cada vez más mutantes de las secuencias maestras dominantes que de las no dominantes lo que finalmente limita la ventaja a cierto nivel ya que la mutación merma la población de las secuencias más frecuentes en mayor cantidad que la de las menos frecuentes.

En los sistemas genéticos reales, ya sean naturales o simulados, los efectos descritos pueden tener gran influencia en los resultados que se obtengan. Por ejemplo, si la tasa de mutación es muy cercana a la  $p^*$ , existe el riesgo de que las secuencias maestras se extingan iniciándose una búsqueda aleatoria (en vez de ordenada) en el dominio de las secuencias genéticas hasta que por azar se llegue nuevamente a obtener una secuencia maestra, la cual a su vez, corre el riesgo de extinguirse nuevamente. Si por el contrario, la tasa de mutación  $p$  es muy baja, el sistema se encuentra en un proceso muy largo de cambios además de limitarse mucho el grado de exploración alrededor de las secuencias maestras reduciéndose la probabilidad de encontrar mutantes lejanas con mejor aptitud que las que se tienen en población.

Una posibilidad para que la evolución mantenga el orden (hasta lo posible) sería que al inicio la tasa de mutación  $p$  fuera muy baja para que las secuencias maestras ejercieran rápidamente su dominio sobre las no maestras. Luego, si la  $p$  aumentara gradualmente, se podría llegar en pocas generaciones al punto de equilibrio manteniéndose siempre el dominio de las secuencias maestras al máximo (reduciéndose el riesgo de su extinción) y al mismo tiempo se iría aumentando el grado de exploración alrededor de éstas.

### 1.4.1.3 Poblaciones con genotipos de longitudes diferentes

Si se considera ahora una población cuyos individuos pueden tener diferentes longitudes en sus genotipos, el comportamiento de la población no puede describirse mediante la ecuación (1.1) y debe usarse el siguiente sistema de ecuaciones:

$$\begin{aligned}
 P_i^{(1)}(t+1) &= \sum_{j=1}^{2^1} p^{2_{ij}} (1-p)^{1-2_{ij}} \frac{f_j}{\bar{f}(t)} P_j^{(1)}(t) \\
 P_i^{(2)}(t+1) &= \sum_{j=1}^{2^2} p^{2_{ij}} (1-p)^{2-2_{ij}} \frac{f_j}{\bar{f}(t)} P_j^{(2)}(t) \\
 &\gg \\
 P_i^{(N)}(t+1) &= \sum_{j=1}^{2^N} p^{2_{ij}} (1-p)^{N-2_{ij}} \frac{f_j}{\bar{f}(t)} P_j^{(N)}(t)
 \end{aligned} \tag{1.20}$$

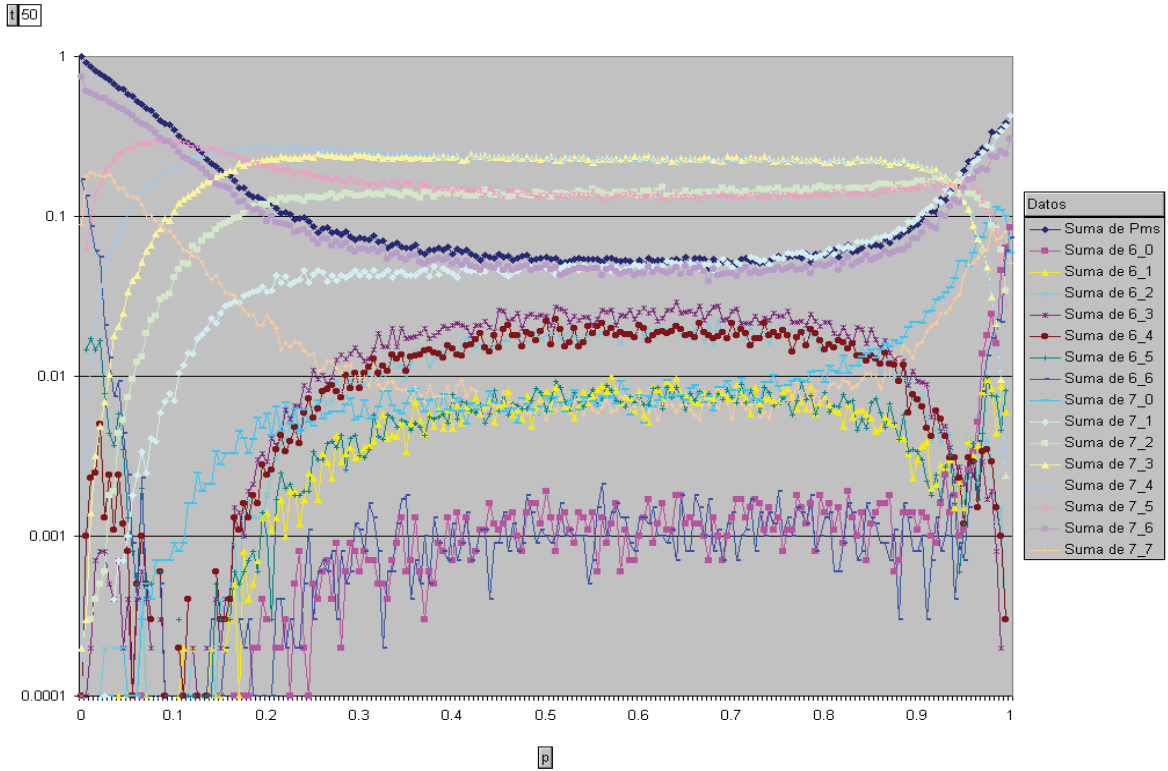
donde  $P_i^{(n)}$  es la fracción de la población que ocupan los individuos cuyas secuencias son de longitud  $n$  y del tipo  $i$ .

El sistema (1.20) puede escribirse como



$$P_i^{(n)}(t) = \sum_{j=1}^{2^n} p^{2_{ij}} (1-p)^{N-2_{ij}} \frac{f_j}{f(t)} P_j^{(n)}(t) \quad 1 \leq n \leq N$$

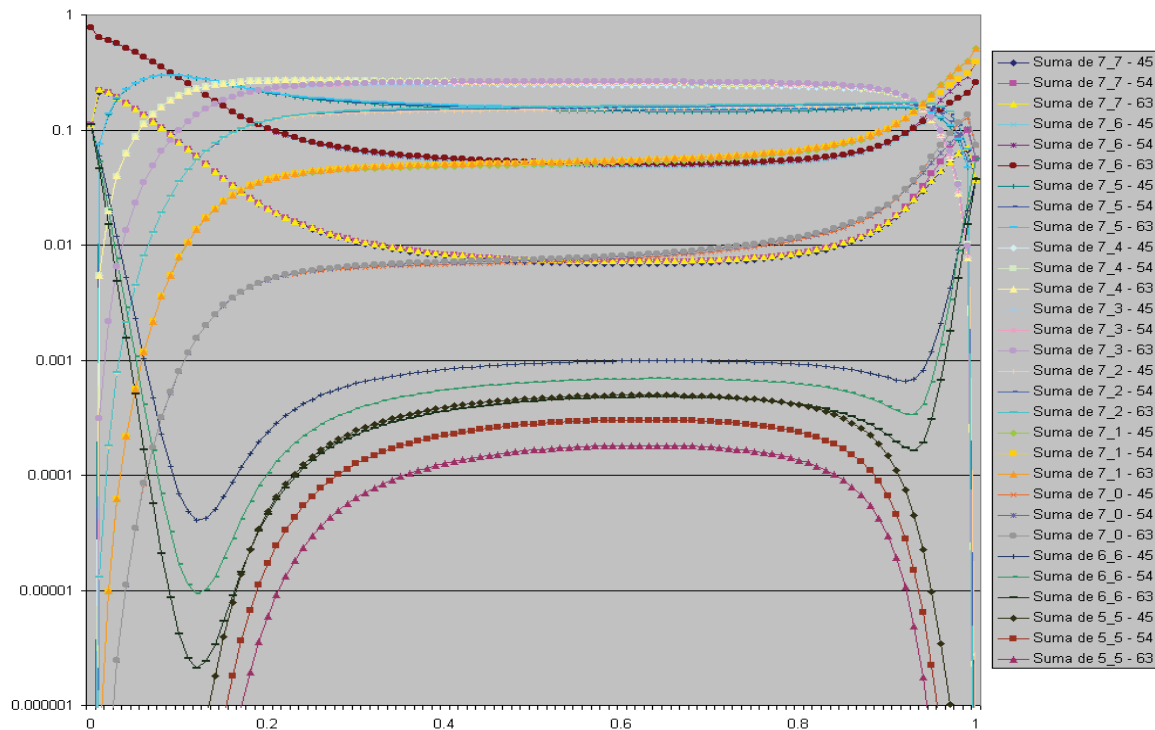
En este modelo se considera que la mutación solamente produce cambios en el valor de los genes de las secuencias y no en la longitud de las mismas. Por esto, puede pensarse que cada conjunto de individuos con genotipos de la misma longitud forman un sistema genético que compite contra los otros sistemas por formar parte de una población que se conserva constante en el número de individuos que la constituyen.



**Figura 1-29** Concentraciones en un sistema genético con diferentes longitudes en los genotipos de sus individuos. Inicialmente se tienen 10,000 individuos con genotipos cuyas longitudes van desde 1 hasta 7 genes. Las secuencias maestras se consideran aquellas que tengan por lo menos 6 unos. Se muestran las concentraciones después de 50 generaciones. La leyenda 6\_0 indica secuencias de longitud 6 con cero unos.

En la **Figura 1-29** se muestran los resultados de un experimento con población finita de 10,000 individuos con secuencias de longitudes desde 1 hasta 7 genes considerando a las que tienen 6 o más genes en 1 ( $L = 6$ ) como las secuencias maestras e iniciando desde una población uniformemente distribuida. Se observa que todas las secuencias de longitud 6 tienen concentraciones muy bajas. La secuencia maestra de longitud 6 tiene una concentración máxima cuando  $p = 0$  y otro máximo local en  $p = 1$ . Esto indica que la selección nunca favorece a las secuencias de longitud 6 incluso cuando estas son secuencias maestras a menos que  $p$  sea 0 o 1 que es cuando el efecto de la población inicial no se pierde.

En la **Figura 1-30** se muestra el comportamiento del mismo sistema en tres generaciones distintas (45, 54 y 63) en el límite de una población infinita para lo cual se integraron las ecuaciones (1.20). Se observa que todas las secuencias con longitudes menores a 7 (no todas se muestran) tienden a reducir su concentración de manera exponencial siendo dominadas totalmente por las secuencias de longitud 7 (la máxima) cuyas concentraciones están ya prácticamente en estado estacionario en estas generaciones. Esto puede explicarse en dos partes: Primero, usando el concepto de “aptitud efectiva” [ 60 ], se tiene que la secuencia maestra de longitud 7 con 7 genes en 1 no pierde tanta aptitud efectiva cuando hay mutación ya que sus vecinas son también secuencias maestras, mientras que las secuencias maestras de longitud 7 con 6 genes en 1 pierden más aptitud efectiva. La secuencia maestra de longitud 6 con 6 genes en 1 pierde aún más aptitud efectiva y por eso sus curvas son más bajas aunque están ya en estado estable. Segundo, las secuencias maestras de longitud 7 con 6 genes en 1 se mantienen estables ya que son “alimentadas” por las secuencias maestras de longitud 7 con 7 genes en 1, cosa que no sucede con las secuencias maestras de longitud 6 ya que la mutación, en este caso, no puede hacer que las secuencias cambien de longitud. El efecto final es como si se tuviera un conjunto de secuencias (las de longitud 7) con cierta aptitud efectiva mayor a la aptitud efectiva de otro conjunto (las de longitud 6) con una tasa de mutación 0 entre ellos produciéndose un dominio total del primer conjunto sobre el segundo.

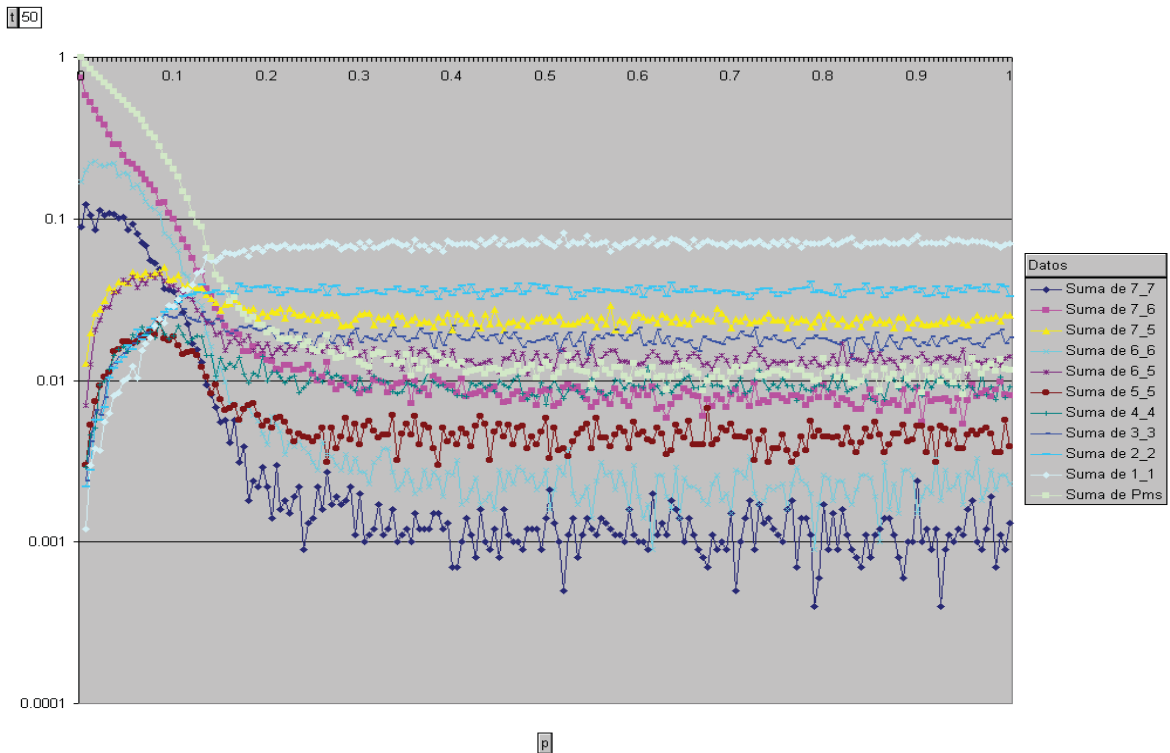


**Figura 1-30** Concentraciones obtenidas a partir de la integración de las ecuaciones (1.20) para  $N = 7, L = 6, G = 2$  en las generaciones 45, 54 y 63. Sólo se muestran las secuencias de longitud 6 con 6 genes en 1 y las de longitud 5 con 5 genes en 1 ya que las demás tienen concentraciones aún menores que éstas.

#### 1.4.1.4 Mutación capaz de modificar la longitud de los genomas

Para extender el análisis hasta aquí realizado a la programación genética (PG) y a otros entornos donde los genotipos codifican diferentes objetos como árboles, se introduce la mutación “*headless chicken*” o pollo descabezado [ 34 ]. Este tipo de mutación es frecuente en la PG en donde al mutar un nodo de un árbol se hace mutar a éste y a todos los nodos que dependen de él reemplazándose todo el subárbol debajo de ese nodo por otro generado al azar.

Una mutación *headless chicken* sobre secuencias genéticas lineales (como las que se han usado antes en este documento) no sólo debería afectar al gen en donde se presentó sino también a otros que dependen de él. Se determinó para los experimentos de esta sección que, a partir del gen en donde se dé la mutación, el resto de la secuencia genética sea reemplazado por otra secuencia aleatoria de longitud aleatoria pero sin exceder el máximo número de genes permitido ( $N$ ) en la secuencia resultante. Primero se elige una longitud (para el resto de la secuencia) entre las posibles longitudes con igual probabilidad y luego se eligen los valores de cada gen también con igual probabilidad (incluyendo al gen en donde se dio la mutación).



**Figura 1-31** Concentraciones en un sistema genético de 10,000 individuos con genotipos de longitudes entre 1 y 7 genes,  $L = 6$ ,  $G = 2$  con mutación de *pollo descabezado* después de 50 generaciones.

Una consecuencia directa del uso de este tipo de mutación es que las secuencias pueden mutar hacia secuencias de otra longitud, cosa que no es posible en la sección 1.4.1.3. Además todas las secuencias, de longitud mayor a cero, pueden mutar hacia ellas mismas o hacia cualquier otra secuencia excepto a la de longitud cero. Bajo este modelo, las secuencias de longitud cero

no pueden mutar. Como las secuencias de longitud cero nunca son secuencias maestras, éstas tienden a extinguirse de todos modos en todos los casos.

En la **Figura 1-31** se observa el comportamiento de las concentraciones en la población en un sistema genético con 10,000 individuos bajo este tipo de mutación. La longitud máxima  $N$  de las secuencias es 7 y las que tienen 6 o más unos se consideran las secuencias maestras ( $L = 6$ ) con una  $G = 2$ . Se observa, a diferencia de la **Figura 1-30**, que la concentración de las secuencias de longitudes menores a 7 ya no tienden a extinguirse sino que se estabilizan en una cantidad (a partir de un número de generaciones) aunque siguen siendo las más largas (de longitud 7) las que dominan. Se observa también la existencia de una probabilidad de mutación crítica a partir de la cual la población es prácticamente uniforme. Nótese que para valores de  $p$  cercanos a 1 no se presenta dominio de las secuencias maestras ni de su opuesta ya que las mutaciones no son en este caso hacia la secuencia opuesta sino hacia cualquier secuencia lo que significa que los cambios genéticos son siempre aleatorios.

Obsérvese que, para tasas de mutación cercanas a 0, las secuencias más favorecidas son las de mayor aptitud, es decir, todas las secuencias maestras que son: las de longitud 7 con 7 ó 6 bits en 1 y las de longitud 6 con 6 bits en 1. Las demás son dominadas por éstas, formándose una Cuasiespecie alrededor de las secuencias maestras. Nótese que el fenómeno de la tasa crítica de mutación sigue presente bajo este tipo de mutación.

Bajo la mutación *headless chicken*, la probabilidad de que una secuencia  $j$  de longitud  $n$  mute en otra secuencia  $i$  de longitud  $m$  es

$$(j_{\gamma_n} \rightarrow i_{\gamma_m}) = (1-p)^n \sum_{k=0}^{\min(n-1, m-1, k)} \binom{n-k}{k} p^k (1-p)^{n-k} \frac{1}{N-h} \frac{1}{2^{m-h}} \quad (1.21)$$

donde  $k$  es el número de genes iguales consecutivos al inicio de  $i$  y  $j$  y  $\delta_{ab}$  es el operador de Kroneker que cumple con

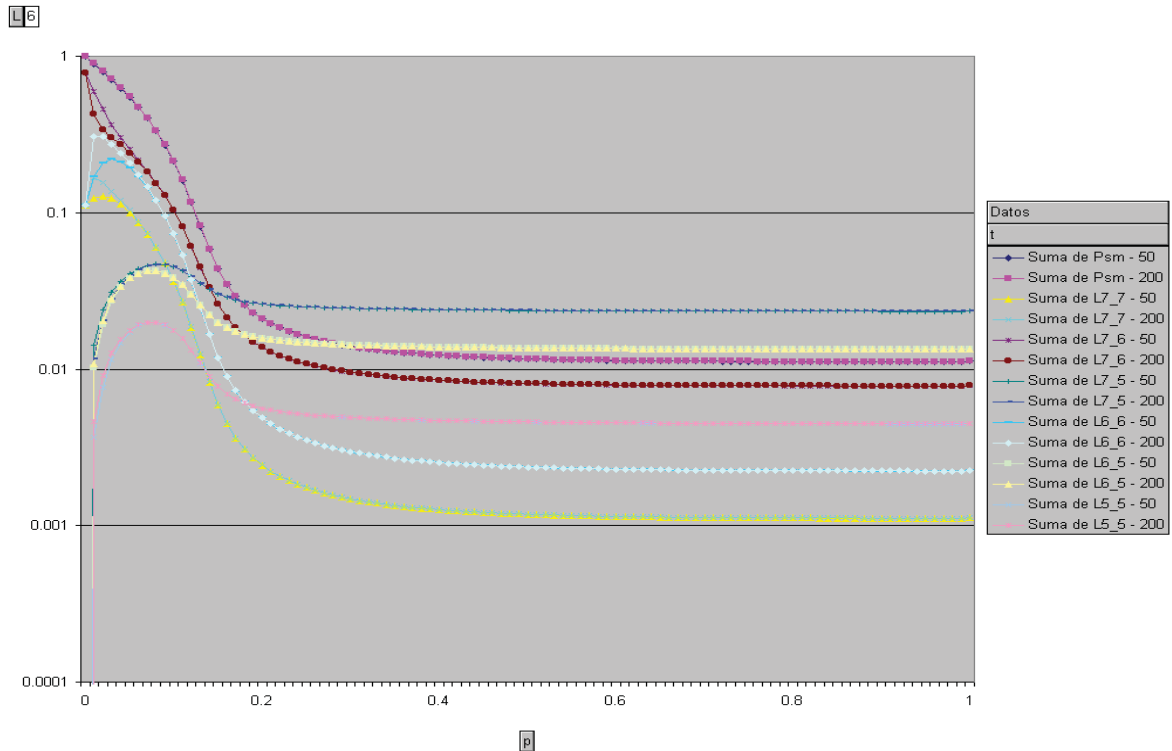
$$\delta_{ab} = \begin{cases} 1 & , a = b \\ 0 & , a \neq b \end{cases}$$

El primer término del lado derecho de (1.21) indica la probabilidad de que no haya mutación siendo  $i$  y  $j$  idénticas. El segundo término indica la probabilidad de que los primeros  $h$  genes no muten  $(1-p)^h$  y que luego haya una mutación (con probabilidad  $p$ ) que “convierta” el resto de la secuencia en una secuencia de longitud tal que la secuencia resultante tenga la misma longitud que  $i$  (con probabilidad  $1/(N-h)$ ) y que sus genes coincidan con los del resto de  $i$  (con probabilidad  $1/2^{m-h}$ ).

La ecuación de evolución para un sistema genético con mutación *headless chicken* sería

$$P_i^{(n)}(t+1) = \sum_{j=1}^N \sum_{k=1}^{2^k} (j \rightarrow i) \frac{f_j}{f(t)} P_j^{(k)}(t)$$

Integrando esta ecuación se obtienen curvas suaves más fáciles de analizar. Los resultados se muestran en la **Figura 1-32** para  $L = 6$  para secuencias maestras y algunas no maestras en las generaciones 50 y 200 para mostrar la tendencia que llevan o si para estas generaciones ya están en estado estable.



**Figura 1-32** Integración de la ecuación de evolución bajo el efecto de la mutación de *pollo descabezado* con  $N = 7$ ,  $L = 6$ ,  $G = 2$  después de 50 y 200 generaciones. La leyenda L7\_7 significa longitud 7 con 7 unos.  $P_{sm}$  es la suma de todas las secuencias maestras.

#### 1.4.2 Degeneración con secuencias maestras separadas

En esta sección se analiza el comportamiento de un sistema genético con selección y mutación en un paisaje de aptitud con varias secuencias maestras separadas, es decir, a más de una mutación de distancia entre ellas. El objetivo es mostrar el comportamiento del sistema en función de la distribución de las secuencias maestras.

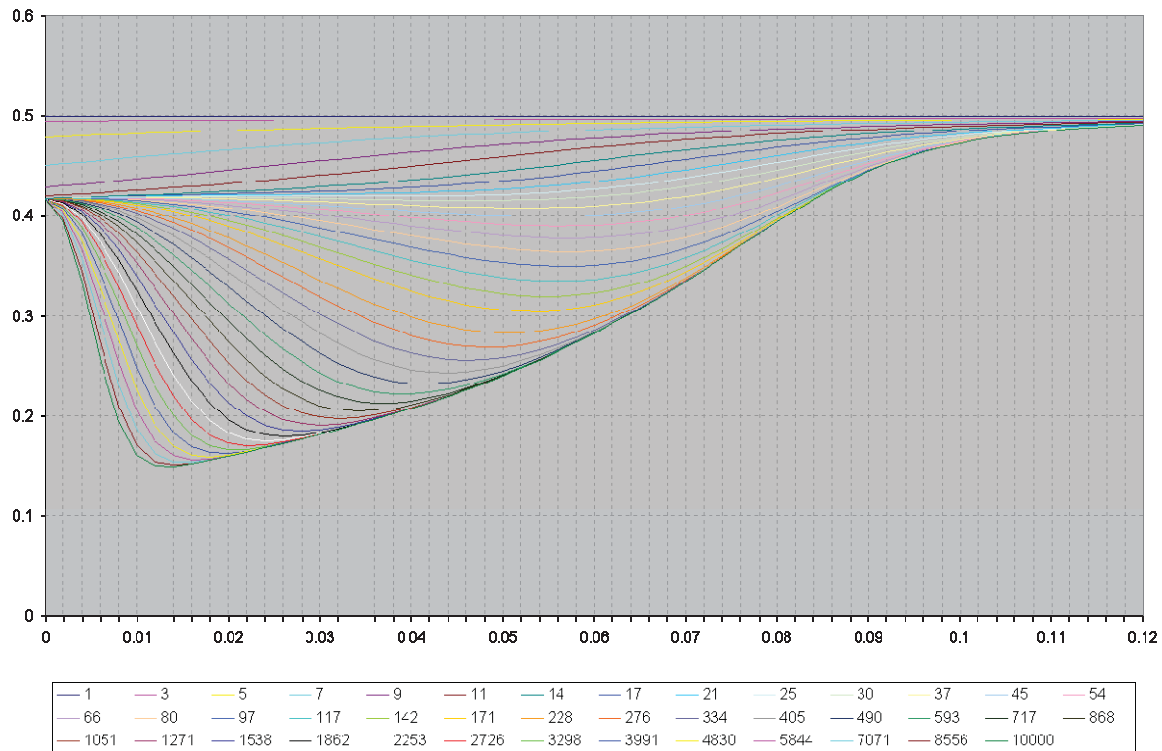
El caso más simple es cuando se tienen 2 agujas aisladas. En este caso simplemente ambas agujas dominan de igual manera y ninguna domina a la otra independientemente de la distancia entre ellas. Un caso más interesante es cuando se tienen 3 agujas. Dependiendo de las distancias de Hamming entre ellas, podrá dominar alguna de ellas. Para medir los resultados se usó la métrica del parámetro de orden el cual está definido de la siguiente manera:

$$orden = \frac{1}{N} \sum_{i=1}^N \sum_{r=1}^8 P_i(t) \delta_{i_r, opt_r}$$

donde  $opt$  es alguna secuencia que generalmente es la secuencia del óptimo,  $N$  es la longitud del genoma,  $i_r$  es el  $r$ -ésimo bit de la secuencia  $i$  y  $\delta$  es el operador de Kronecker que cumple con

$$\delta_a^b = \begin{cases} 1 & , a = b \\ 0 & , a \neq b \end{cases}$$

Este parámetro de orden es el promedio de coincidencias que hay en la población con los bits de un genotipo dado. En otras palabras, esta métrica dice la concentración de unos en la población. Cuando es cero significa que no hay unos y cuando es uno significa que todos los bits en la población son unos.



**Figura 1-33** Parámetro de orden como función de  $p$  para varios valores de  $t$  (generaciones) para el caso de tres agujas: 11111111, 00000011 y 00000000 con  $G = 2$ .

En la **Figura 1-33** se muestran los resultados de un sistema de secuencias de 8 bits con 3 agujas:  $sm_1 = 11111111$ ,  $sm_2 = 00000011$  y  $sm_3 = 00000000$  cuyas aptitudes son el doble de la aptitud de las demás secuencias. Las distancias de Hamming entre estas agujas son:  $\delta_{sm_1, sm_2} = 6$ ,  $\delta_{sm_1, sm_3} = 8$ ,  $\delta_{sm_2, sm_3} = 2$ . Obsérvese como es que para  $p = 0$  son justamente estas tres agujas las que dominan de manera absoluta a toda la población y el parámetro de orden



corresponde a la concentración de unos en estas tres secuencias, es decir, 0.41666. Para valores de  $p$  más allá de la  $p^*$  se tiene una población casi uniforme por lo que el parámetro de orden es muy cercano a 0.5 todo el tiempo. Para otros valores de  $p$  el parámetro de orden tiende hacia un valor menor a 0.5 al paso de las generaciones. Esto se debe a que las secuencias  $sm_2$  y  $sm_3$  que están más juntas una de la otra están dominando a la secuencia  $sm_1$  que está más alejada, ya que el número de mutantes que recibe  $sm_2$  de  $sm_1$  y  $sm_3$  y también el número de mutantes que recibe  $sm_3$  de  $sm_1$  y  $sm_2$  es mayor al número de mutantes que recibe  $sm_1$  de  $sm_2$  y  $sm_3$ .

En paisajes con más agujas aisladas se espera que la secuencia dominante sea aquella que reciba mayor cooperación de sus vecinas. Esta cooperación depende de la aptitud de cada secuencia, de la distancia de Hamming entre las secuencias y de la tasa de mutación  $p$ . En la siguiente sección se estudian casos donde las secuencias “maestras” tienen diferentes aptitudes, es decir, que en el paisaje de aptitud se tienen varias señales de diferentes alturas y anchuras.

## 1.5 Competencia entre Cuasiespecies

Un caso muy común en los sistemas genéticos reales es que en el paisaje de aptitud se tengan varios picos o señales con diferentes alturas y diferentes grados de degeneración compitiendo por atraer a la población hacia ellos. Como se verá en seguida, no en todas las situaciones es la señal más alta la que domina en la población sino que dependiendo de la tasa de mutación y de la forma (altura y anchura) de cada señal será una u otra la que domine. Se da entonces en ocasiones un fenómeno conocido como *transición de fase* en función de la tasa de mutación en el que con cambios muy pequeños en  $p$  se produce un cambio abrupto en el comportamiento del sistema.

Las transiciones de fase proporcionan un marco formal a partir del cual se pueden entender los efectos que tienen los operadores en un sistema genético. Las transiciones de fase se pueden definir como el cambio cualitativo en el comportamiento de un sistema provocado por modificaciones leves en los parámetros del mismo. Estas pueden ser de *segundo* o de *primer* orden. Las de *segundo* orden son aquellas en las que la derivada de la función presenta saltos abruptos en sus valores en función de la variable que se esté considerando. Las de *primer* orden presentan cambios abruptos en la función misma.

### 1.5.1 Transiciones de fase de segundo orden

En el caso de un sistema genético evolucionando bajo selección y mutación, se presentan transiciones de fase de segundo orden respecto a la tasa de mutación  $p$ . Véase en la **Figura 1-4** de la página 37 cómo es que, para valores de  $p$  intermedios, la forma de la Cuasiespecie es prácticamente plana y constante respecto a  $p$ , mientras que para valores de  $p$  cercanos a cero y cercanos a uno se tiene que esta forma depende del valor de  $p$ . Estos cambios en el comportamiento del sistema respecto a  $p$  son transiciones continuas, es decir, que la forma de

la Cuasiespecie pasa de un comportamiento a otro de manera continua y por lo tanto se considera una transición de fase de segundo orden.

La transición de fase mencionada es muy marcada en el paisaje de aptitud de aguja en un pajar con  $N$  grande. En otros paisajes puede ser más suave e inclusive dejar de ser una verdadera transición de fase. Esto puede verse por ejemplo en la **Figura 1-17** en donde se tienen paisajes con varias agujas juntas que, cuando son demasiadas agujas, la transición de fase se desvanece y solo queda un tipo de comportamiento del sistema respecto a  $p$  (véase también la **Figura 1-21**).

Esta no es la única transición de fase posible en la Cuasiespecie. También se pueden dar transiciones de fase de primer orden cuando existen dos o más señales de diferente forma y tamaño en las que se pasa del dominio de una al dominio de la otra con cambios muy pequeños en  $p$ . En seguida se estudian este tipo de casos.

### 1.5.2 **Transiciones de fase de primer orden**

Este tipo de transición de fase se distingue del anterior por presentar discontinuidades en las concentraciones respecto a alguna de las variables. Las transiciones de fase de primer orden pueden presentarse en sistemas genéticos evolucionando bajo selección y mutación sólo en casos especiales de paisajes de aptitud. Dichos casos son aquellos en donde se tiene más de un valor máximo local en el paisaje de aptitud y que además cumplen ciertas condiciones.

Se ha visto ya que una señal más ancha (degenerada) será más robusta ante la mutación que una delgada. Entonces es claro que si en el paisaje de aptitud hay dos señales de la misma altura pero de distinta anchura la que debe dominar es la ancha. También es claro que si se tienen dos señales igual de anchas pero de diferente altura será la más alta la que domine. La situación menos obvia es cuando hay una señal alta pero delgada compitiendo contra una ancha pero más baja. En este caso, la señal que domina depende de la aptitud efectiva de las señales la cual depende de la tasa de mutación que se use.

Para analizar esto se usará primero un sistema genético con genomas de longitud 20 y un paisaje de aptitud en el que la clase 20 tiene aptitud 2, las clases 0 y 1 tienen aptitud 1.9 y las demás 1. Esto genera un paisaje con 2 máximos locales: el de la clase 20 y el de las clases 0 y 1. El primero es más alto y el segundo más ancho (véase **Figura 1-34**).

La secuencia dominante no siempre es la de mayor aptitud (la clase 20) [ 19 ]. Dependiendo de la tasa de mutación, la secuencia dominante puede ser la clase 20 o la clase 0. Para valores bajos de mutación la secuencia dominante es la de mayor aptitud, la clase 20, y para valores mayores a cierta tasa  $p_f$ , que se denominará de *transición*, la secuencia dominante es la clase 0. En la **Figura 1-35** se grafican las concentraciones de cada clase de *Hamming* como función de  $p$  en fase estable. Se nota un cambio abrupto en las concentraciones cuando  $p$  es aproximadamente 0.11. Ésta es una transición de primer orden ya que las curvas presentan una discontinuidad. Es notable en la gráfica que, a pesar de que en las secuencias no maestras no hay ninguna diferencia en su valor de aptitud, unas tienen mayor concentración que otras. Esta diferencia se debe a que unas reciben mayor cantidad de mutantes desde la secuencia



dominante que otras ya que están más cerca de ella de la misma manera que se vio en el paisaje de aguja en un pajar solamente que aquí es la señal dominante la que alimenta a las demás. Se puede decir que se da una competencia entre las Cuasiespecies asociadas con ambas señales y siempre una termina dominando a la otra.

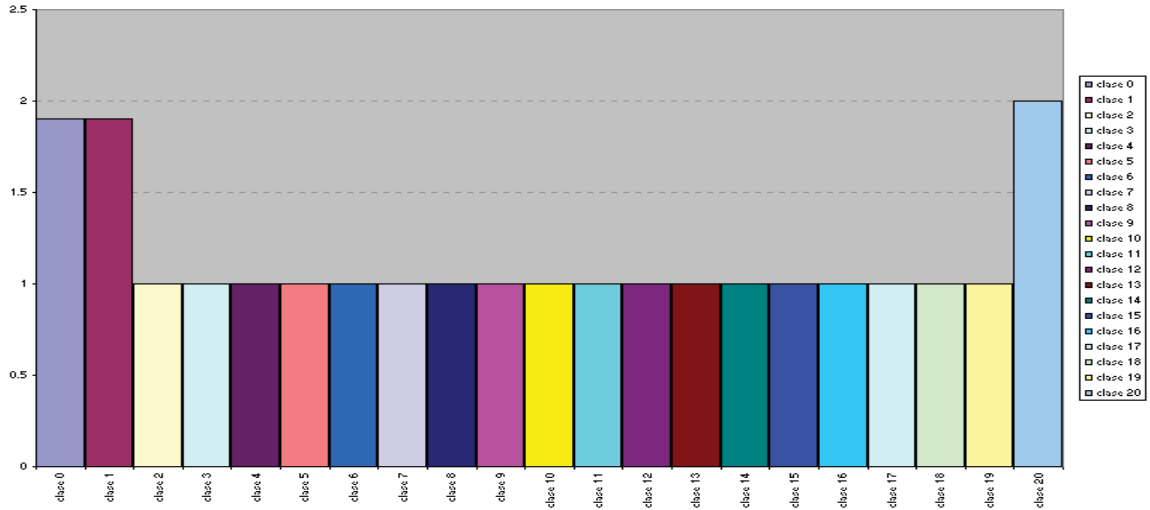


Figura 1-34 Paisaje de aptitud con dos picos. Uno alto y el otro ancho.

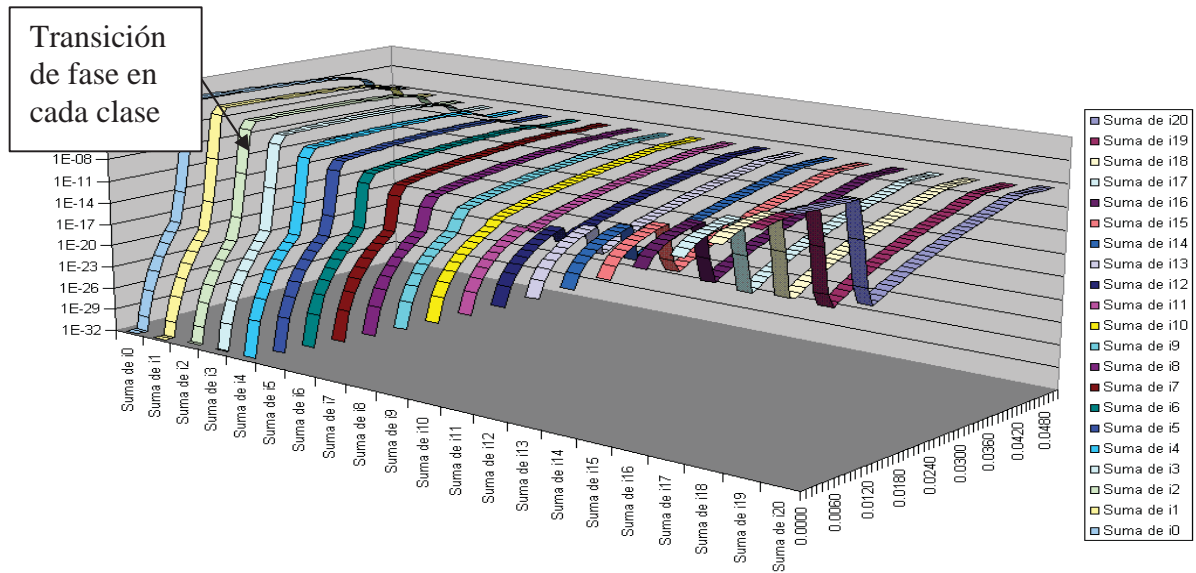
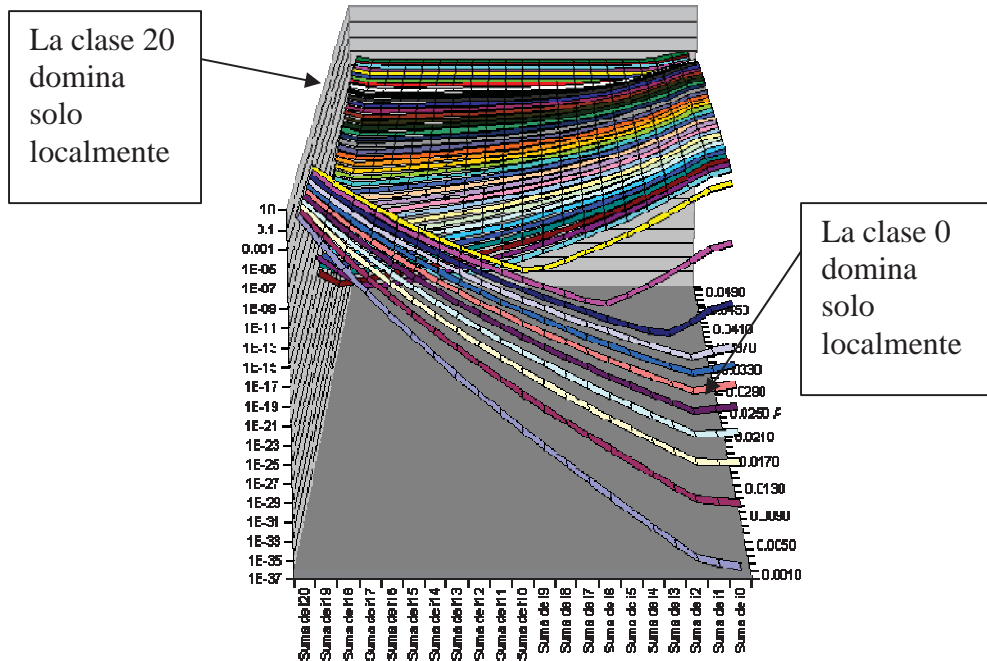


Figura 1-35 Aquí se muestran las concentraciones de cada secuencia dentro de cada clase como función de  $p$  después de 3000 generaciones.



**Figura 1-36** En ésta gráfica se muestra la distribución de la población después de 3000 generaciones para varios valores de  $p$ .

En la **Figura 1-36** se muestra la distribución de la población (en fase estable) con varios valores de  $p$  para mostrar y comparar las concentraciones de las secuencias genéticas o Cuasiespecies dada alguna  $p$  fija. Para valores bajos de  $p$  (al frente), las secuencias de las clases 0 y 1 ejercen un dominio solamente sobre las secuencias más cercanas a ellas. Incluso hay secuencias no maestras que tienen concentraciones mayores a las de éstas. Lo mismo pasa con la clase 20 cuando  $p$  es mayor a la tasa de transición (al fondo). Esto se debe a que la Cuasiespecie asociada al pico dominante aplasta prácticamente a la otra.

En la **Figura 1-37** se muestran las concentraciones acumuladas de cada clase de hamming como función del tiempo (en este caso el pico alto es la clase 0 y el ancho las clases 19 y 20) para una  $p$  menor a la tasa de transición. Nótese que a pesar de que durante una buena parte de la evolución se tiene más concentración de la Cuasiespecie asociada al pico ancho, después empieza a dominar la del pico alto.

La **Figura 1-38** es análoga a la **Figura 1-37** pero a una tasa de mutación mayor a la tasa de transición. Es claro que ahora es la Cuasiespecie del pico ancho la que domina después de algunas generaciones.

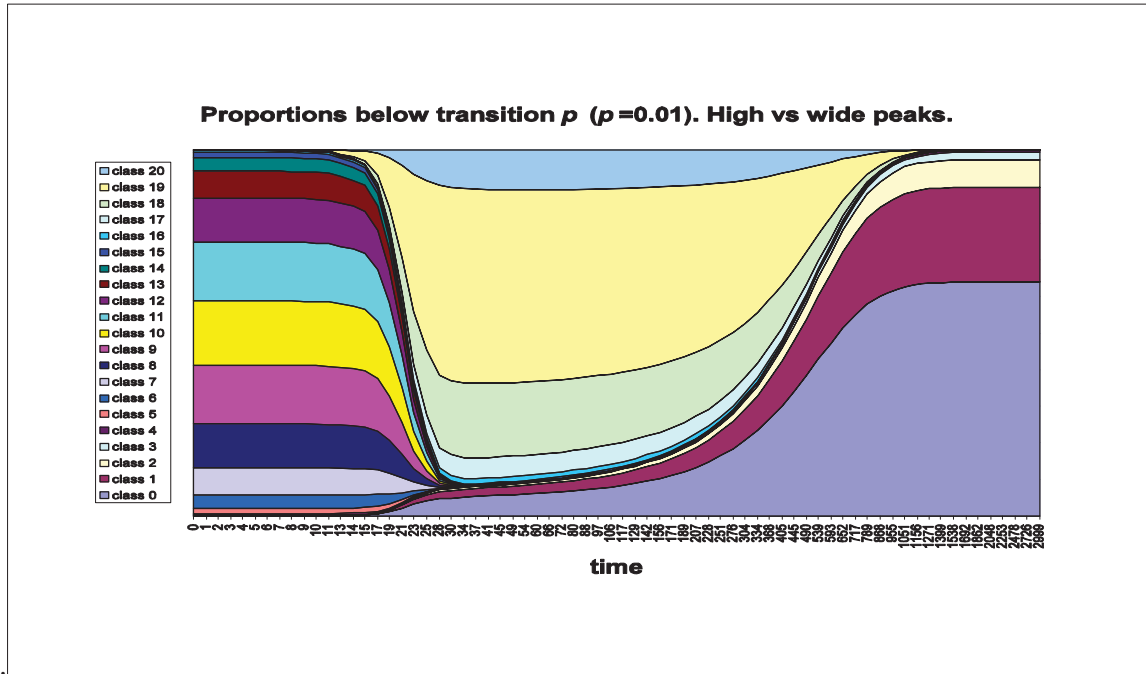


Figura 1-37 Concentraciones de cada clase de Hamming acumuladas en función del tiempo para el paisaje con una aguja alta (en este caso la clase 0) y otra ancha (en este caso las clases 20 y 19) a una tasa de mutación menor a la  $p$  de transición.

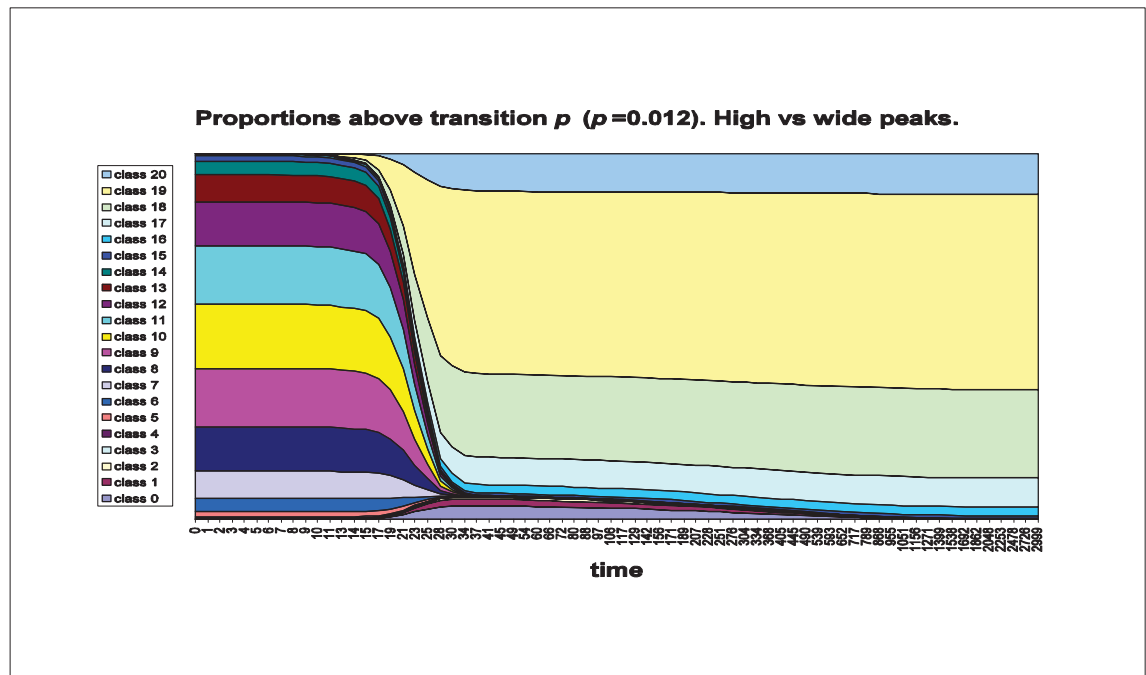


Figura 1-38 Concentraciones de cada clase de Hamming acumuladas en función del tiempo para el paisaje con una aguja alta (en este caso la clase 0) y otra ancha (en este caso las clases 20 y 19) a una tasa de mutación mayor a la  $p$  de transición.

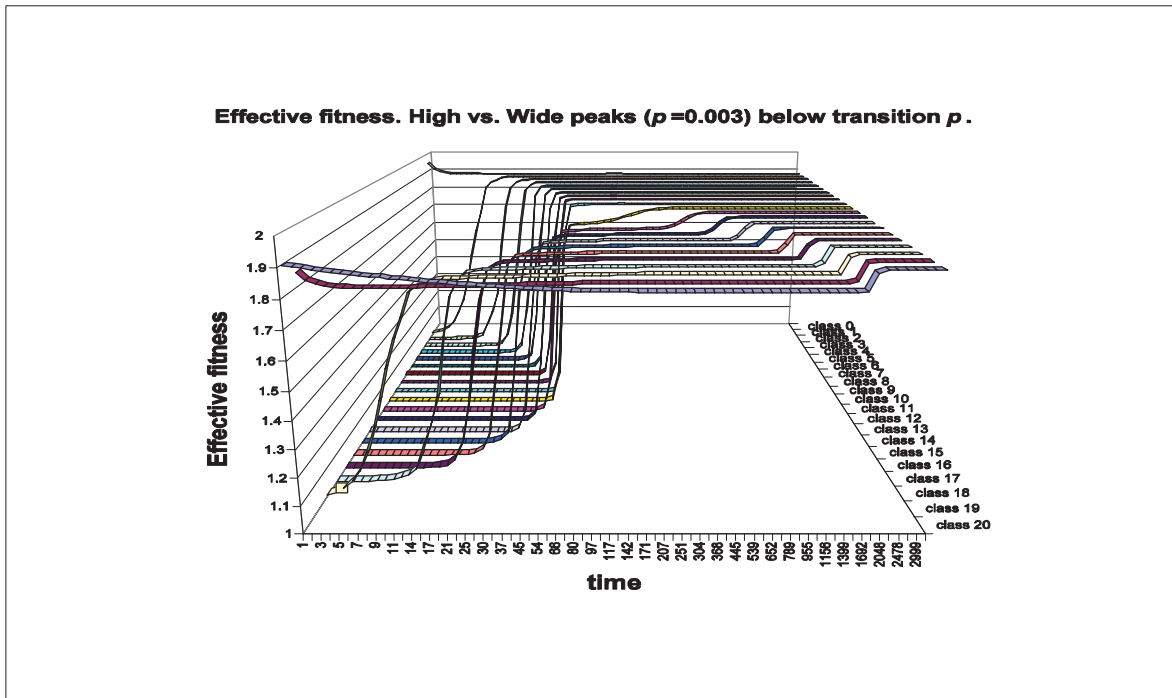


Figura 1-39 Aptitud efectiva de cada clase como función del tiempo en el paisaje de aguja alta (clase 0) y aguja ancha (clases 20 y 19) con una  $p$  menor a la tasa de transición.  $N = 20$ .

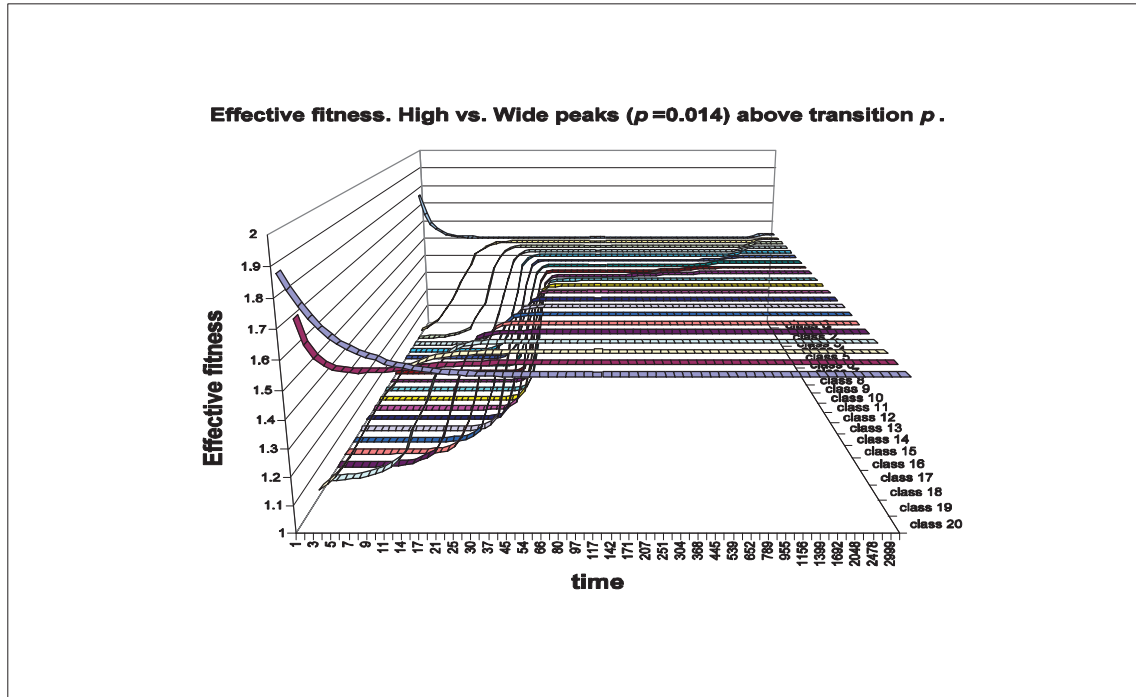


Figura 1-40 Aptitud efectiva de cada clase como función del tiempo en el paisaje de aguja alta (clase 0) y aguja ancha (clases 20 y 19) con una  $p$  mayor a la tasa de transición.  $N = 20$ .

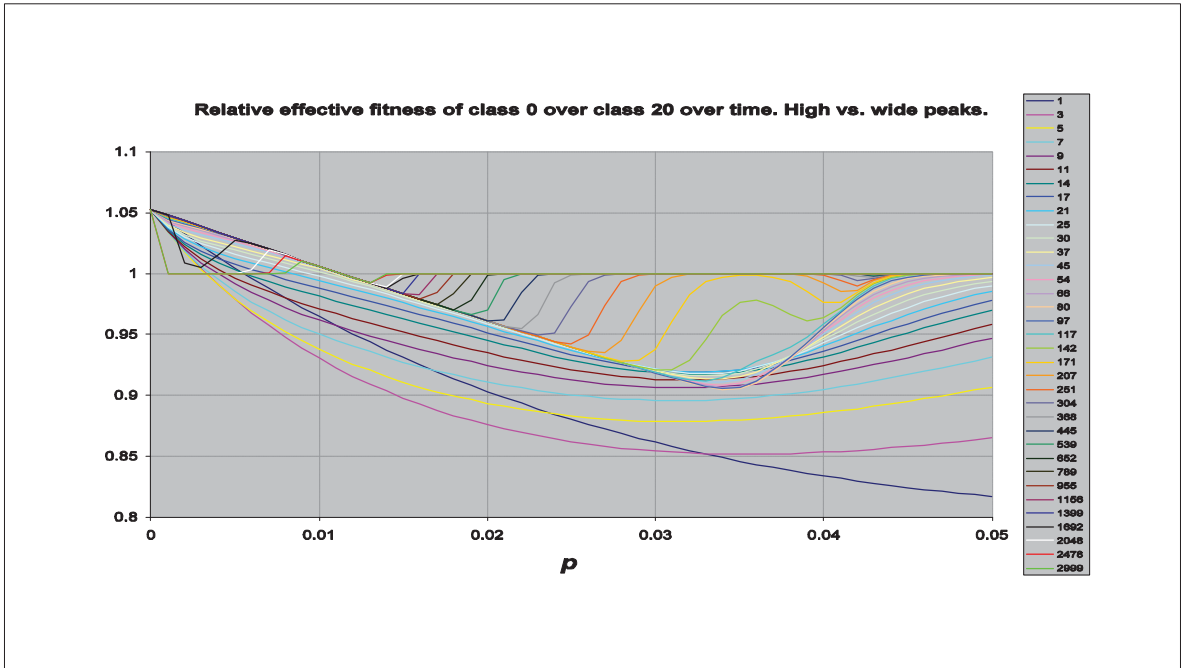


Figura 1-41 Aptitud efectiva relativa de la aguja alta respecto a la aguja ancha en función de  $p$  para diferentes generaciones a partir de una población uniforme infinita.

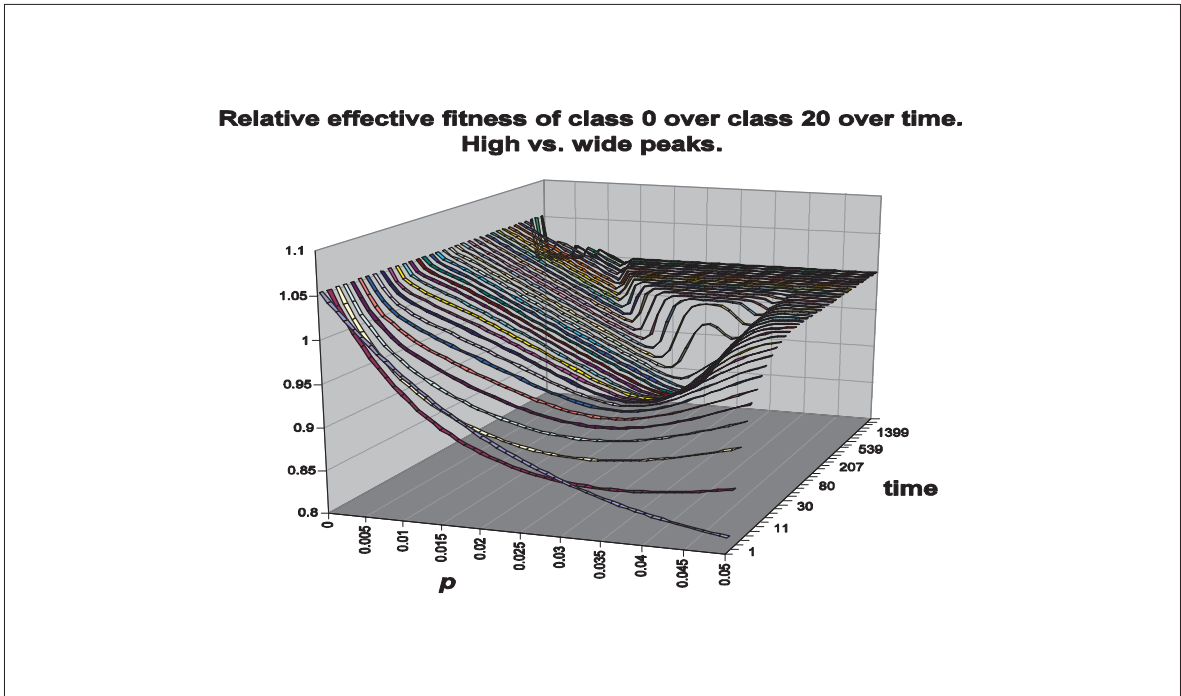


Figura 1-42 Vista en 3D de la aptitud efectiva relativa de la aguja alta respecto a la aguja ancha en función de  $p$  para diferentes generaciones a partir de una población uniforme infinita

Este efecto puede explicarse mediante la aptitud efectiva. Véase en la **Figura 1-39**, en donde se grafica la aptitud efectiva de cada clase de Hamming para una  $p$  menor a la  $p$  de transición  $p_f$ , que la Cuasiespecie del pico alto (al fondo) tiene mayor aptitud efectiva y domina a la del pico ancho (al frente) y en la **Figura 1-40** el caso contrario que es cuando  $p$  es mayor a  $p_f$  en donde es la Cuasiespecie del pico ancho la que domina.

Para medir la ventaja en aptitud efectiva de una secuencia sobre otra se define la *aptitud efectiva relativa* como el cociente de las aptitudes efectivas de ambas dada por

$$f_{eff}^{i,j}(t) = \frac{f_{eff}^i(t)}{f_{eff}^j(t)}$$

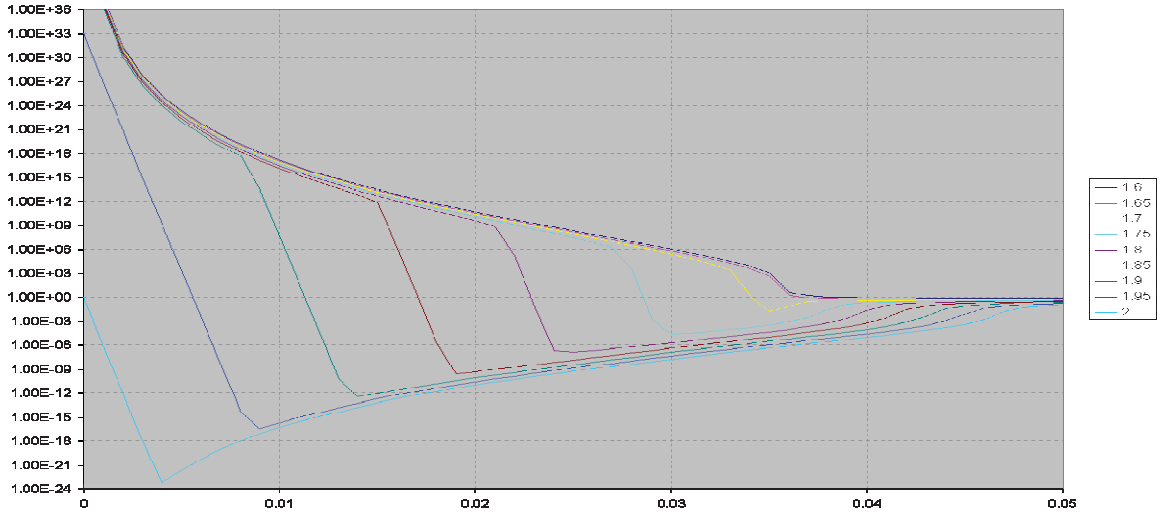
En la **Figura 1-41** y la **Figura 1-42** se muestra la aptitud efectiva relativa de la aguja alta respecto a la ancha en función de  $p$  para varios valores de  $t$ . Los valores de esta métrica mayores a 1 implican dominio de la aguja alta y viceversa.

Destaca aquí que durante la fase transitoria esta aptitud efectiva relativa depende casi linealmente de  $p$  cuando  $p$  es menor a la tasa crítica de ambas agujas lo cual indica que la ventaja relativa, si existe, disminuye directamente con el aumento de  $p$ .

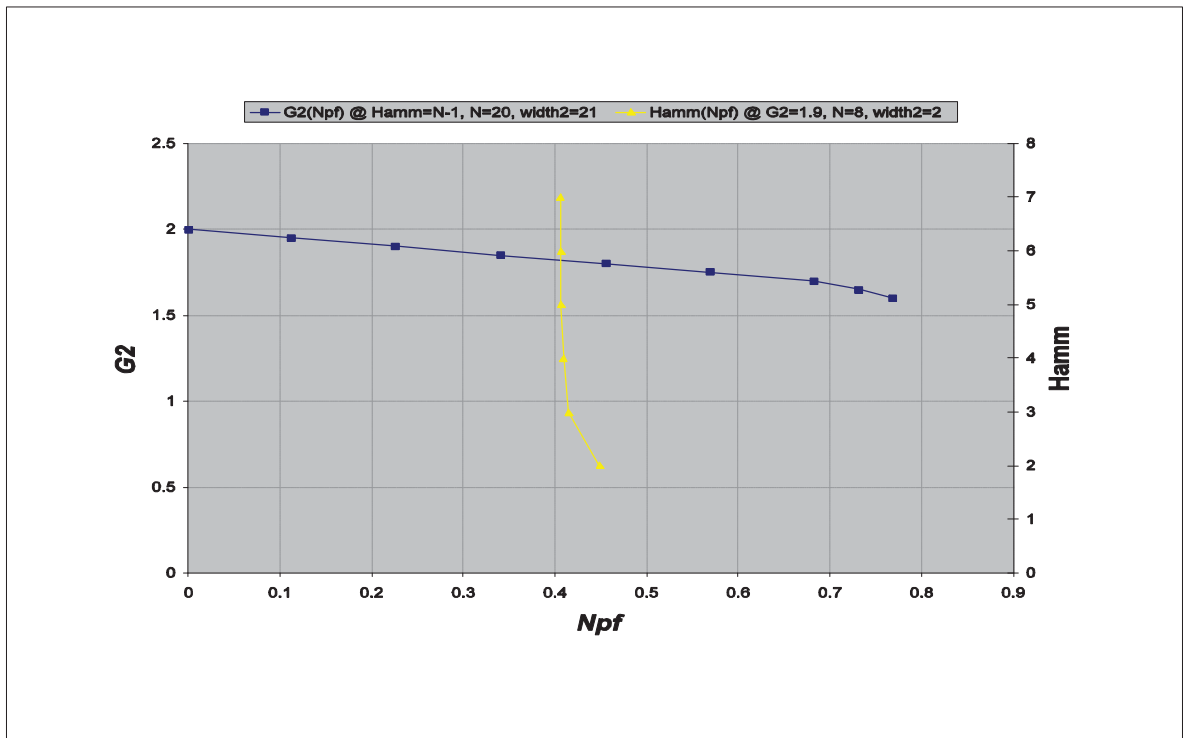
El valor de  $p$  con el cual las curvas cruzan el valor 1 es justamente el valor de la  $p$  de transición  $p_f$ . En la fase estable, esta métrica es siempre 1 indicando que ninguna de las señales ejerce ventaja alguna sobre la otra. Esto no significa que las concentraciones sean iguales sino que se mantienen constantes.

También se puede observar que, para valores de  $p$  cercanos a la  $p$  de transición  $p_f$  la fase transitoria dura considerablemente más tiempo. Esto último se debe a que la aptitud efectiva de ambas señales es muy parecida haciéndose más fuerte la competencia entre ellas. Sin embargo, la señal que tenga ventaja irá dominando a la otra de manera exponencial hasta lograr un dominio total de la otra.

Pero ahora surge la pregunta ¿Cómo depende el valor de la tasa de transición  $p_f$  del paisaje de aptitud, es decir, de la altura y anchura de las señales que están compitiendo? En la **Figura 1-43** se muestra la razón de la concentración de la clase 20 respecto a la clase 0 como función de  $p$  después de 3000 generaciones para distintos valores de la aptitud de las clases 0 y 1. La tasa de transición es aquella en la que las curvas cruzan por 1. De estos datos se obtiene la gráfica de la **Figura 1-44** (en cuadros azules). Se observa un comportamiento casi lineal de  $p_f$  respecto a la aptitud de las clases 0 y 1 mientras dicha tasa no llegue a la  $p$  crítica de la señal alta que en este caso es cuando  $Np_f = \ln(2) = 0.69$ . Cuando  $p_f$  está por encima de este valor significa que la tasa de mutación debe ser mayor a la  $p$  crítica de la señal alta para que la señal ancha domine convirtiéndose ésta efectivamente en una señal de ruido solamente (véase la sección 0 en la que se analizan paisajes con ruido). Cuando la aptitud de las clases 0 y 1 es igual a la de la clase 20, la  $p$  de transición es cero y por lo tanto la clase 20 nunca es dominante.



**Figura 1-43** Relación entre la concentración de la clase 20 (aguja alta) respecto a la clase 0 (contenida en la aguja ancha) como función de  $p$  para distintos valores de aptitud de las clases 0 y 1 (después de 3000 generaciones). La  $p$  de transición es donde las curvas cruzan por 1.



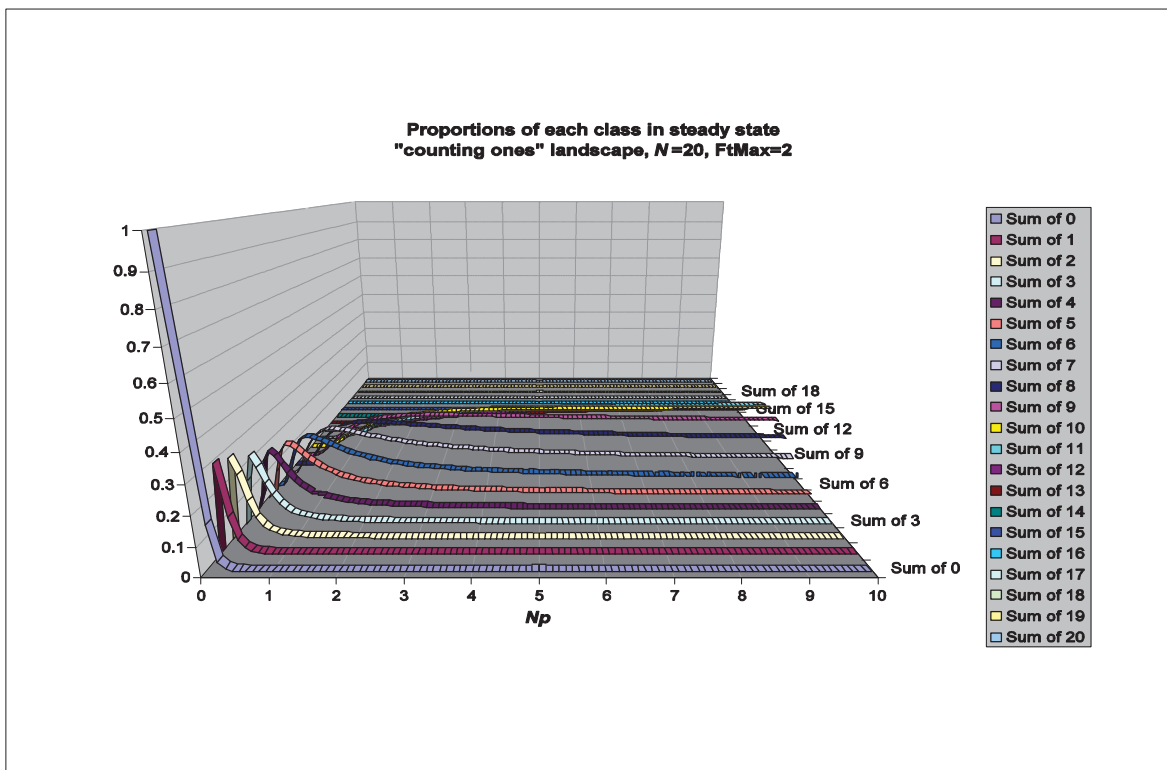
**Figura 1-44** Dependencia de  $p_f$  respecto a la altura de la señal ancha y de la distancia de Hamming entre la señal alta y la ancha.

En la misma **Figura 1-44** se grafica la forma en que  $p_f$  depende de la distancia de Hamming entre las dos señales. Se observa un aumento en  $p_f$  cuando ambas señales están muy cerca indicando claramente que al estar cerca la competencia entre ellas se convierte en cooperación

ya que los mutantes de una alimentan a la otra.

## 1.6 Cuencas de atracción

Una cuenca de atracción es una zona del paisaje de aptitud que atrae la población hacia un punto interior de ésta. Generalmente, la aptitud de los genotipos disminuye como función de su distancia de Hamming de ese punto interior. Un ejemplo simple es el paisaje en el que la aptitud está dada por el número de unos de los genotipos, en donde todo el paisaje es una cuenca de atracción hacia el punto donde el genotipo tiene solamente unos. En la literatura se conoce a este paisaje como ONEMAX porque tiene solo un máximo local o *counting ones* (contando unos). Bajo este paisaje y con  $N = 20$ , la integración de la ecuación de evolución hasta llegar a un punto cerca del punto de equilibrio o estado estacionario, da como resultado las curvas de la **Figura 1-45** asociadas por clase de Hamming desde el óptimo que es la clase 0.

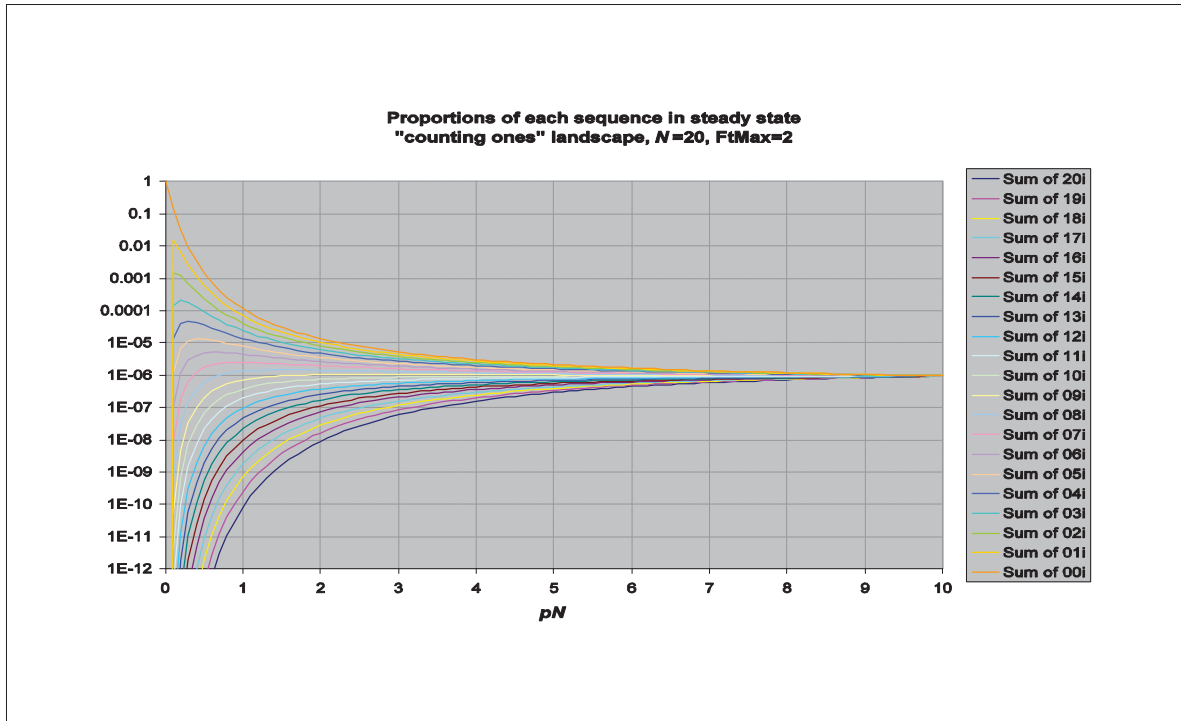


**Figura 1-45** Concentraciones por clase de Hamming en función de  $p$  en fase estable.

Obsérvese cómo es que cada clase tiene una tasa crítica por encima de la cual su concentración disminuye notablemente. Además, la clase de Hamming dominante cambia conforme la tasa de mutación crece. Esto puede entenderse si se piensa que cada clase de Hamming es una señal degenerada. Por ejemplo, la clase 0 está formada por un solo genotipo, la clase 1 es una señal formada por  $N$  genotipos, la clase 2 por  ${}^N C_2$  genotipos, la 3 por  ${}^N C_3$ , etc. Entonces cada señal tiene una resistencia distinta a la mutación siendo las más robustas las que tienen mayor



grado de degeneración que en este caso es la clase 10. A tasas de mutación bajas domina la clase con mayor aptitud y a tasas altas la de mayor robustez.



**Figura 1-46** Concentraciones de la Figura 1-45 normalizadas por el número de secuencias pertenecientes a cada clase de Hamming.

Lo anterior no implica que el óptimo sea más débil que otros genotipos individuales. La interpretación correcta es que, dada una tasa fija  $p$ , resulta más probable que se tenga en la población alguna secuencia perteneciente a una clase que a otra.

Véase en la **Figura 1-46**, en donde se grafican las mismas concentraciones de cada clase pero ahora normalizadas por el número de genotipos pertenecientes a la clase, que el óptimo es la secuencia más esperada para toda  $p$ . Esto puede explicarse en los términos usados anteriormente cuando dos señales están situadas a una distancia de Hamming 1 una de la otra. En estos casos la señal más baja (sin importar si es más robusta) siempre coopera con la señal más alta con cualquier  $p$ . Se puede pensar que este paisaje está formado por una secuencia de señales contiguas con aptitud decreciente a una distancia de Hamming 1 entre ellas. Así, las señales más bajas cooperan con las más altas y nunca hay competencia entre ninguna de ellas.

En este paisaje se forma rápidamente una cuasiespecie centrada en el óptimo a cualquier tasa de mutación diferente de 0 y de 1. La distribución de los genotipos en la cuasiespecie es la que se ve en la **Figura 1-46**. Es claro que la concentración de la población alrededor del óptimo siempre es mayor en los genotipos cercanos al óptimo y decrece con la distancia de Hamming. En este caso no hay una tasa crítica de mutación debido a la cooperación entre todos los genotipos hacia el óptimo.

## 1.7 Cuasiespecies con población finita

En las secciones anteriores se utilizó el modelo matemático de un sistema genético que evoluciona mediante selección y mutación en el límite de una población infinita en todas las situaciones básicas que dan un panorama general de la dinámica del sistema identificando sus grados de libertad efectivos. Ahora se hará un análisis práctico con poblaciones finitas para corroborar lo anterior y para ver cuales características se modifican y si hay algunas que surgen a partir de este hecho.

Concretamente, se analiza la forma en que la tasa de mutación y el tamaño de la población afectan la búsqueda en términos del esfuerzo necesario para hallar la secuencia óptima (número de evaluaciones de la aptitud de los individuos) y, una vez encontrada, la probabilidad de supervivencia de la secuencia óptima en diferentes paisajes de aptitud.

La idea es encontrar tanto la tasa de mutación como el tamaño de población ideales para ambas situaciones lo cual no siempre es posible lograr y resulta ser en general un compromiso entre ambas.

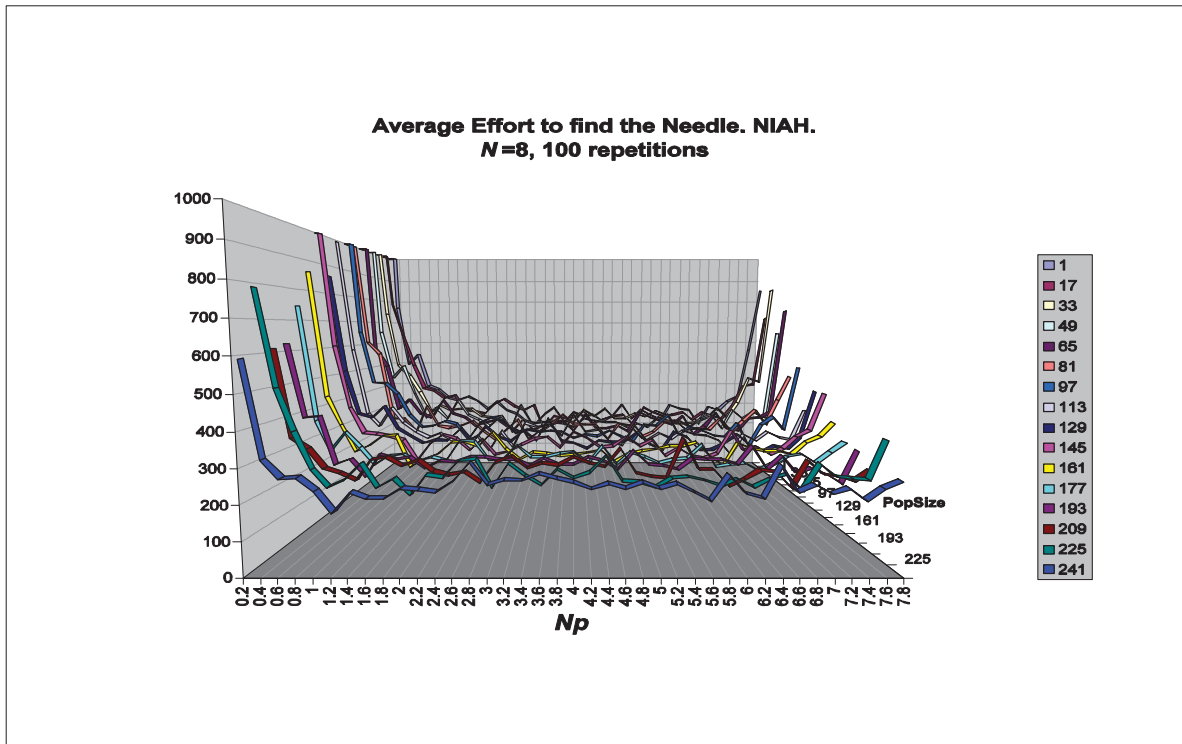
### 1.7.1 *Búsqueda de una aguja en un pajar*

El análisis en esta sección es importante ya que da una visión de la manera óptima de hallar, con una población finita, una señal que no tiene ninguna correlación con el resto del paisaje y de la manera óptima también en que, una vez hallada dicha señal, esta población finita debe aglutinarse para seguir explorando alrededor de ella sin perderla.

#### 1.7.1.1 *Tasa de mutación óptima para encontrar el óptimo*

Cuando no se tiene el óptimo (la aguja) en la población, este paisaje se convierte en un paisaje plano ya que ninguna secuencia no óptima es preferida sobre otra. Esto se traduce en un comportamiento aleatorio de los cambios genéticos en la población y por lo tanto la búsqueda es básicamente aleatoria. Aún así, hay una tasa de mutación que hace la búsqueda más eficiente.

En la **Figura 1-47** se grafica el esfuerzo promedio sobre 100 repeticiones para encontrar el óptimo como función de  $Np$  para diferentes tamaños de población para  $N = 8$ . Es claro que la búsqueda es menos eficiente cuando  $p$  está cerca de 0 y cuando está cerca de 1 ( $Np = 8$ ). La tasa óptima es 0.5 ya que con esta tasa los genotipos tienen una máxima variación lográndose una exploración del espacio más completa. Con cualquier otro valor de  $p$  se tendría mayor probabilidad de que un individuo visite lugares que ya visitó anteriormente siendo necesarias más evaluaciones de la función objetivo.



**Figura 1-47** Promedio sobre 100 repeticiones del esfuerzo de búsqueda en el paisaje de aguja en un pajar en función de la tasa de mutación  $p$  (escalada por  $N$ ) con diferentes tamaños de población.  $N = 8$ .

#### 1.7.1.2 Tasa de mutación y tamaño de la población óptimas para maximizar la probabilidad de supervivencia de la secuencia maestra

Al usar poblaciones finitas, la dinámica del sistema cambia ya que se introduce una componente aleatoria. Esto se refleja en la concentración de la secuencia maestra o aguja como se ve en la **Figura 1-48** en donde se hizo correr el algoritmo con diferentes tamaños de población iniciando siempre con la aguja presente cuya  $G = 2$ ,  $N = 30$ . A menor tamaño de la población se tiene una menor concentración promedio de la secuencia maestra, especialmente si la tasa de mutación es cercana a la tasa crítica teórica, lo que hace que efectivamente la tasa crítica se reduzca. Pero ¿a qué se debe esto? En la sección 0 se pudo intuir que una tasa de mutación muy cercana a la tasa crítica teórica no sería muy buena ya que el dominio que ejerce la secuencia maestra (o las secuencias que forman la señal en el paisaje de aptitud) no puede ser rápido, así que habrá una mayor probabilidad de que, debido a mutaciones, se extinga dicha secuencia y por consiguiente no se dirija la búsqueda hacia ella sino que la población se “pierda” haciendo búsqueda aleatoria ya que el resto del paisaje es plano.

Para corroborar lo anterior, se define ahora una métrica para calcular la probabilidad de supervivencia de la secuencia maestra una vez que se le ha encontrado mediante la búsqueda genética. Dada una población aleatoria pero con uno y solo un individuo con la genética de la secuencia maestra se deja correr el algoritmo un número de generaciones y se calcula la fracción del número de repeticiones en que dicha secuencia sigue presente en la población

para diferentes tasas de mutación.

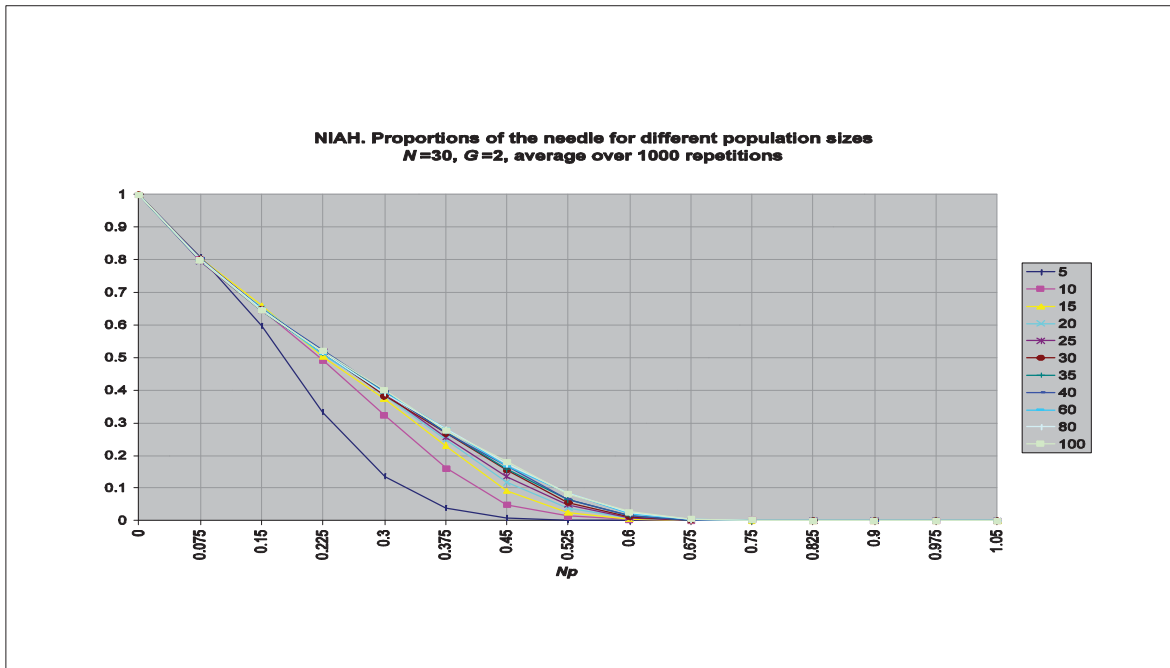


Figura 1-48 Concentraciones promedio sobre 1000 repeticiones de la aguja con diferentes tamaños de población.

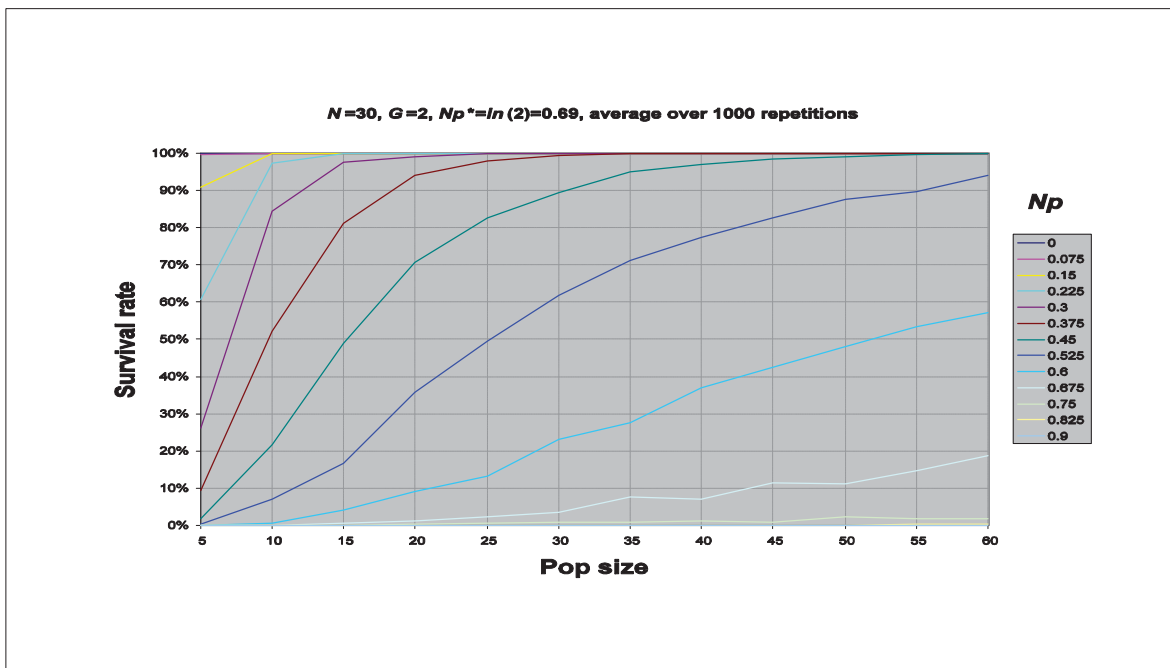


Figura 1-49 Probabilidad de supervivencia de la secuencia maestra una vez encontrada.

En la **Figura 1-49** se grafica esta probabilidad de supervivencia de la secuencia maestra como función del tamaño de la población  $s$  con 1000 repeticiones para diferentes tasas de mutación

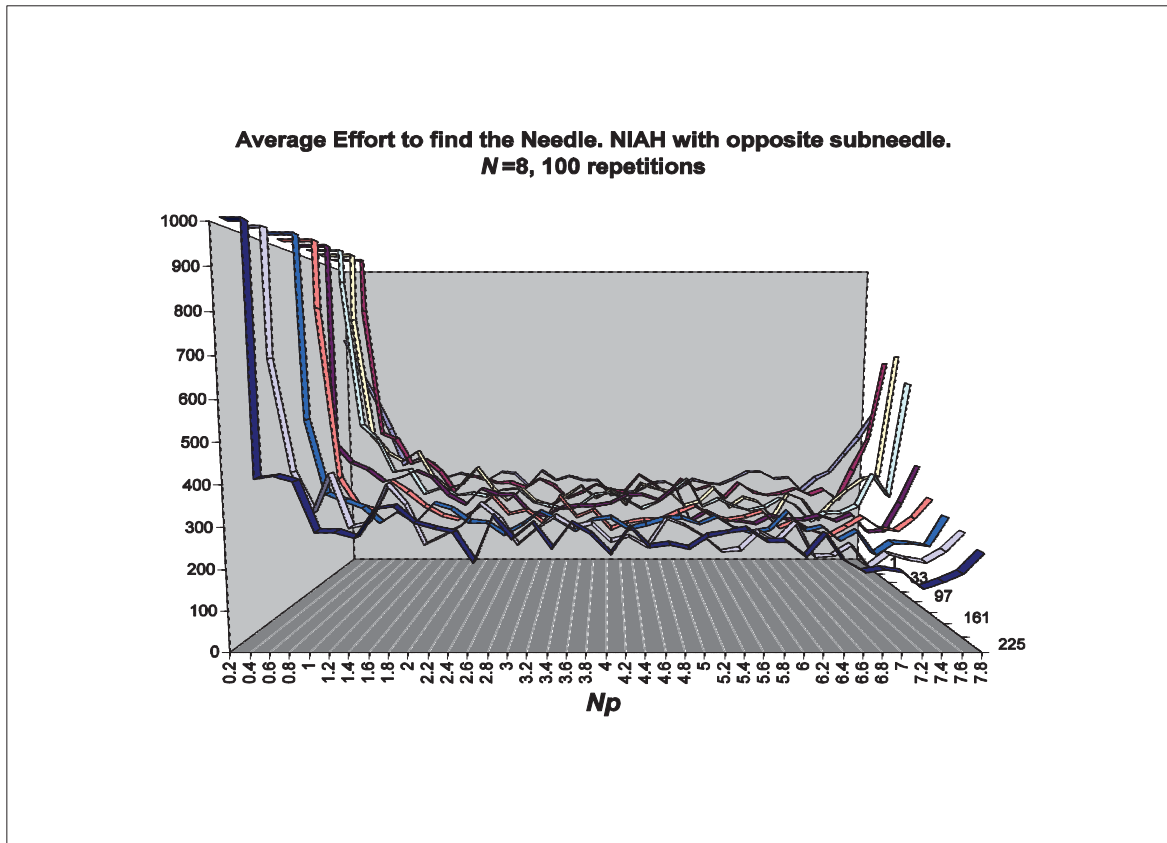
$p$  en el rango de 0 a algo un poco mayor a la tasa crítica teórica  $p^*$  que en este caso es  $\ln(2)/30 = 0.0231$  (en la gráfica  $p$  está escalada con  $N$  así que  $Np^*$  es  $\ln(2) = 0.69$ ). De la figura es claro que mientras más grande es  $s$  hay mayor probabilidad de supervivencia de la secuencia maestra para todas las  $p$ . Por supuesto, en el límite de población infinita todas las curvas tienden a 1 pero lo importante es que algunas se acercan a 1 desde valores bajos de  $s$  y son justamente las  $p$  más bajas. Entonces, el mejor valor de  $p$  para conservar el óptimo es 0 y el mejor valor de  $s$  es infinito. Estos valores son claramente malos para la búsqueda genética en general ya que sin mutación no hay exploración y las poblaciones muy grandes son computacionalmente imposibles. Se requiere entonces un buen compromiso entre explotación y exploración además de eficiencia. Si se requiere una probabilidad de supervivencia mínima por ejemplo de 90%, cada valor fijo de  $p$  implica que se debe tener una  $s$  mínima para cumplirla. Si al mismo tiempo se requiere una  $p$  alta para tener mucha exploración, se necesitará una población sumamente alta para lograr ese 90%. Por esto, debe elegirse una tasa de mutación lo más alta posible pero cuya curva tenga una pendiente relativamente alta respecto al tamaño de la población. Así con  $s$  baja podrán alcanzarse probabilidades de supervivencia altas. Esa tasa óptima es mucho menor a  $p^*$  y puede aproximarse como  $p^*/2$ . Con esta tasa, un buen valor de  $s$  es aproximadamente igual a  $N$ .

### 1.7.2 **Búsqueda en paisajes con dos señales**

Cuando hay dos señales en el paisaje donde una es más alta (el óptimo) que la otra se puede dar uno de dos casos: que la señal del óptimo sea hallada primero en cuyo caso el comportamiento del sistema (al menos la fase de búsqueda del óptimo) es el mismo que cuando se tiene solo una señal, o que la otra señal sea hallada primero con la consecuencia de que la población formará una Cuasiespecie alrededor de ésta afectando la manera en que se sigue explorando el espacio en búsqueda del óptimo.

#### 1.7.2.1 *Subóptimo anticorrelacionado*

En el paisaje definido en la **Figura 1-34** se tienen 2 señales opuestas o anticorrelacionadas en el espacio de búsqueda. Los resultados de la eficiencia de búsqueda (esfuerzo) están en la **Figura 1-50** para diferentes tamaños de población y en función de  $Np$  (el esfuerzo máximo permitido fue de 1000 evaluaciones). Bajo este paisaje la búsqueda es óptima cuando la tasa de mutación  $p$  está cerca de 1 ya que, en caso de que se encuentre primero la señal opuesta al óptimo, ésta atraerá la población hacia ella y en la siguiente generación esta atracción será “enviada” por mutación hacia el óptimo aumentándose la probabilidad de encontrarlo. Este caso demuestra que la tasa de mutación ideal puede ser algo sumamente distante de lo que normalmente se usa que son valores cercanos a cero.



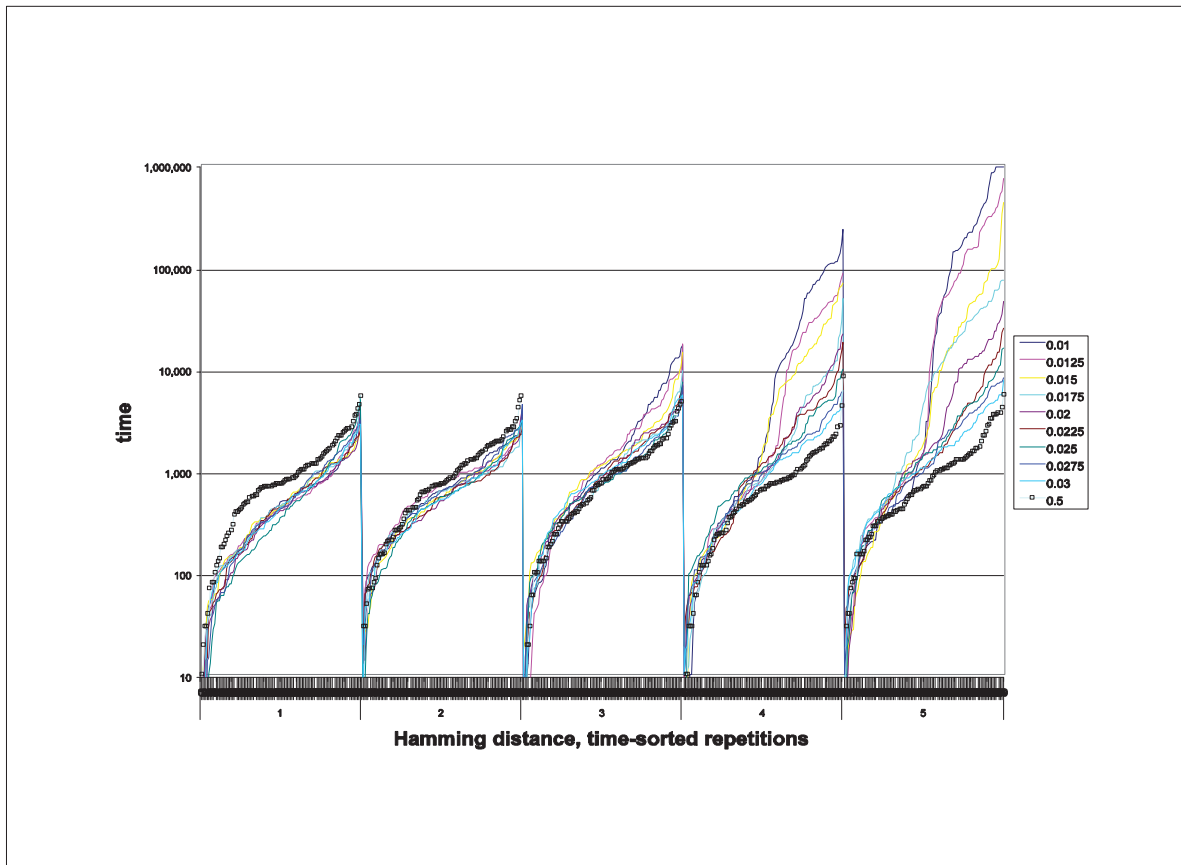
**Figura 1-50** Esfuerzo de búsqueda en un paisaje con una señal anticorrelacionada con la señal del óptimo.

### 1.7.2.2 Cooperación o decepción

Ya se vio que cuando la señal del subóptimo está anticorrelacionada con la del óptimo resulta posible que su presencia se traduzca en un impedimento para la búsqueda al menos con tasas de mutación menores a  $\frac{1}{2}$ . Pero ¿Qué pasa si esta señal está cerca del óptimo? Si esta señal es hallada primero, la población será atraída hacia ella y consecuentemente también hacia el óptimo por lo que se espera que, en lugar de estorbar, la señal del subóptimo ayude a encontrar el óptimo. Para ver esto se trazan en la **Figura 1-51** las curvas características de búsqueda (ver descripción en el Apéndice 2) para diferentes tasas de mutación y para diferentes distancias de Hamming entre las señales a partir de una población inicial aleatoria de 10 individuos con  $N = 10$  en donde la señal del subóptimo tiene un ancho de un solo genotipo y una aptitud del doble que el resto de los genotipos excepto el óptimo.

Se tiene que, si la señal del subóptimo está colocada a una distancia de Hamming baja de la señal del óptimo, la búsqueda (en el rango de  $p$  entre 0 y algo superior a la tasa crítica de la señal del subóptimo) es más eficiente que la búsqueda aleatoria (con  $p = \frac{1}{2}$  representada en cuadros negros) ya que, si se encuentra primero el subóptimo, éste atrae la búsqueda cerca del óptimo (a tasas de mutación menores que  $\frac{1}{2}$ ). Al aumentar la distancia de Hamming, la cooperación que presentaba el subóptimo se va convirtiendo en competencia llegando al punto

en que la búsqueda ya no puede ser mejor que búsqueda aleatoria. En otras palabras, una señal adicional en el paisaje de aptitud ayuda a la búsqueda si está cerca del óptimo o correlacionado a éste y estorba si está lejos o anticorrelacionado con el óptimo. Nótese en la **Figura 1-51** que a distancias de Hamming altas (derecha) la búsqueda se ve afectada solamente en la mitad de las repeticiones. Esto es porque ambas señales tienen el mismo ancho, es decir, están formadas por un solo genotipo y por lo tanto tienen la misma probabilidad de ser encontradas primero. Si la señal del subóptimo fuera más robusta habría mayor probabilidad de encontrarle primero que a la señal del óptimo lo cual afectaría a un número mayor de repeticiones.



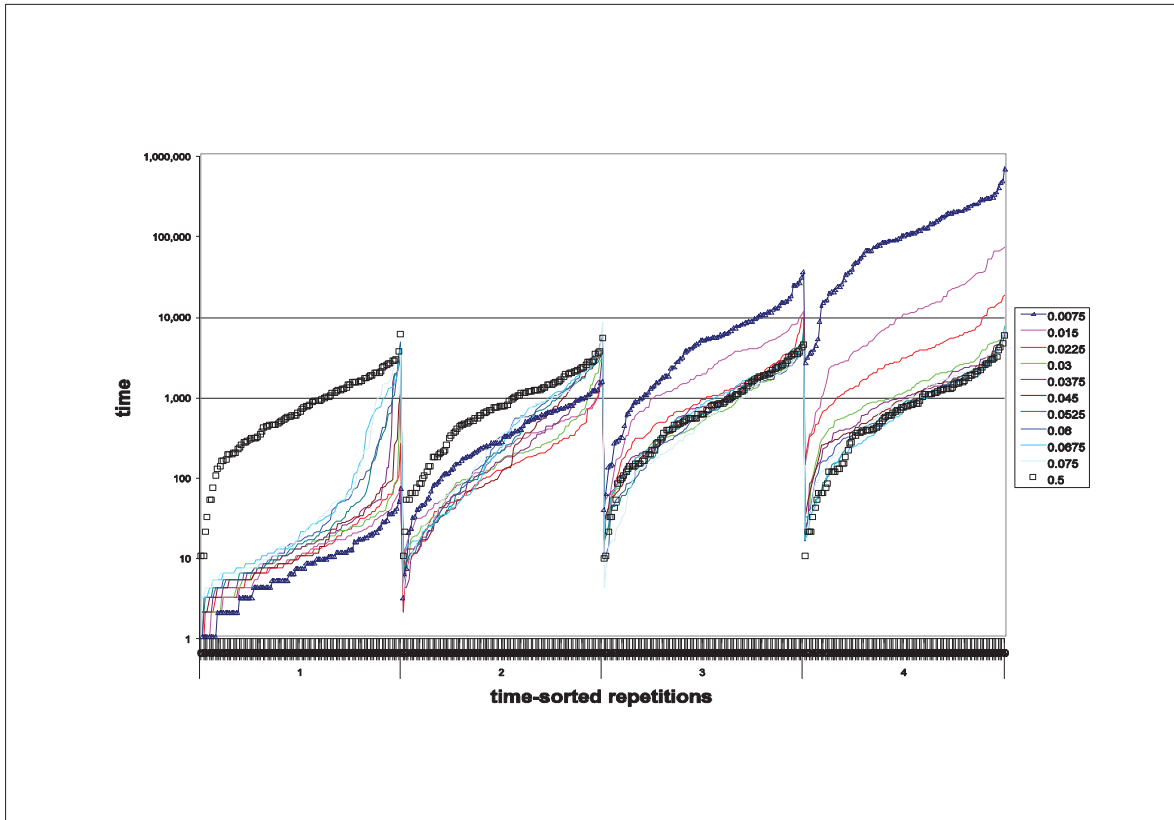
**Figura 1-51** Curvas características de búsqueda para diferentes tasas de mutación en un paisaje con un subóptimo a diferentes distancias de hamming del óptimo.  $N = 10$ .

En la misma **Figura 1-51** puede verse que, a una distancia de Hamming 1, prácticamente todas las tasas de mutación usadas (menores a 0.5) resultan igual de eficientes. Para distancias de Hamming altas, la tasa de mutación óptima es la más alta del rango ya que, de ser encontrado el subóptimo primero, sería necesario que la población no se acercara a éste siendo necesaria una tasa de mutación mayor a la tasa de mutación crítica del subóptimo para así poder escapar de su influencia. Entonces, la tasa ideal que cubre todos los casos es alta y la eficiencia de la búsqueda depende de la distancia de Hamming. En el mejor de los casos la búsqueda es un poco mejor que la búsqueda aleatoria.



1.7.2.3 *Búsqueda de una señal desde la otra*

Suponiendo ahora que la población ya fue atraída hacia el subóptimo, la tasa de mutación ideal sí depende de la distancia de Hamming al óptimo. Este es un caso muy común en los paisajes de aptitud reales en donde la Cuasiespecie presenta saltos puntuales desde un subóptimo hacia otro supuestamente mejor.



**Figura 1-52** Curvas características de búsqueda en un paisaje de aptitud con un pico subóptimo y otro óptimo a diferentes distancias de hamming donde la población inicial está concentrada en el subóptimo.  $N = 10$ .

En la **Figura 1-52** se trazan las curvas características de búsqueda a partir de una población concentrada en el subóptimo el cual está a diferentes distancias de Hamming del óptimo. En esta situación la población debe explorar el espacio alrededor del subóptimo para hallar el óptimo. Cuando el óptimo está a una distancia de Hamming 1, casi cualquier tasa de mutación baja resulta muy buena (bastante mejor que búsqueda aleatoria), siendo la mejor tasa una muy baja pero mayor que 0 ya que esta tasa es obviamente demasiado baja. Al pasar al caso de distancia de Hamming 2 se ve que la tasa de mutación ideal es ahora un poco mayor y la eficiencia ya no es tan diferente a la búsqueda aleatoria. Si se sigue aumentando la distancia se verá un aumento en la  $p$  ideal hasta el punto en el que tal tasa ideal es 0.5 como se ve en el caso de distancia de Hamming 3 y 4. Esto es otra vez porque la búsqueda tiende a concentrarse cerca del subóptimo haciéndose necesario que la población salga de ese subóptimo para tener mayor probabilidad de dar con el óptimo.



## 1.8 Conclusiones

Son varias las lecciones que se deben aprender del presente capítulo respecto a cómo funciona la evolución con selección y mutación. Hay algunas conclusiones a partir del modelo con población infinita y otras a partir de simulaciones con poblaciones finitas usadas como herramienta de búsqueda.

### Población infinita

Cuando la selección es muy fuerte, el único grado de libertad efectivo (GLE) es la tasa de mutación ya que en cada generación la población se reduce a individuos con el genotipo dominante para luego mutar desde ese hacia el resto del espacio de búsqueda. Cuando la selección es débil el sistema presenta un comportamiento más complejo siendo  $G$  (la ganancia de aptitud) un GLE.

En el paisaje de aptitud de aguja en un pajar existe una tasa de mutación crítica  $p^*$ , más allá de la cual la “aguja” o secuencia maestra ya no puede dominar a la población (este es el mismo resultado que el de Eigen y colaboradores [ 19 ]). Se demostró que existe una relación universal entre  $p^*$  y el grado de degeneración de la aguja (o tamaño relativo de su red neutral), en la que el GLE es el porcentaje del espacio de los genotipos que ocupa la señal. Se observa que, más allá del 5% de esta ocupación, el fenómeno de la  $p^*$  desaparece. Se observan también dos fenómenos superpuestos durante la creación de la cuasiespecie: dominio de las secuencias maestras sobre el resto de las secuencias y dominio de unas secuencias maestras sobre otras igualmente aptas pero con diferente aptitud efectiva. Cuando  $p$  está muy cerca de  $p^*$ , el primero se hace más lento y cuando  $p$  está cerca de 0 es el segundo el que se hace más lento.

Si se tienen dos o más señales en el paisaje, se puede dar o no, una transición de fase de primer orden en la distribución final de las concentraciones de las secuencias genéticas respecto a la tasa de mutación (resultado de Eigen y colaboradores [ 19 ]). Se mostró que la existencia y el valor de la tasa de transición  $p_f$  dependen de los detalles de las dos señales como son su altura y grado de degeneración. Dada una tasa de mutación fija, la altura y degeneración de las señales determinan la forma y fuerza de cada cuasiespecie correspondiente y, por lo tanto, también determina cuál de las dos cuasiespecies domina a la otra. La aptitud efectiva de ambas cuasiespecies se debilita al aumentar la tasa de mutación pero en diferentes cantidades. En caso de que la señal alta se debilite más rápidamente y la más baja sea suficientemente fuerte, puede darse el caso de una transición de fase de primer orden, dado que para tasas bajas la aguja alta domina a la baja, pero a tasas mayores puede ser que la aguja alta esté tan debilitada que la baja resulta más fuerte y por lo tanto tiene mayor aptitud efectiva y termina dominando en la población. Cuando la señal ancha es demasiado baja comparada con la alta, se comporta simplemente como una señal de ruido. Esto ocurre cuando  $p_f$  es mayor que la  $p^*$  de la señal alta.

En caso de que las dos señales están cerca, la tasa de transición crece ya que en lugar de que las cuasiespecies compitan, se da una cooperación de parte de la señal más baja hacia la más alta. Cuando están lejos, ambas cuasiespecies evolucionan independientemente terminando

siempre en el dominio de la más fuerte sobre la más débil.

Las cuencas de atracción en el paisaje de aptitud tienen un efecto similar al de una señal robusta. La población tiende a formar una cuasiespecie alrededor del centro de dicha cuenca que es el punto más alto de ella.

En general puede decirse que el GLE en sistemas evolutivos mediante selección y mutación es la cuasiespecie. Todas las señales en el paisaje de aptitud generan una cuasiespecie alrededor de ellas. Dicha señal puede ser una aguja aislada o una cuenca de atracción. La fuerza de una señal la determinan su ganancia de aptitud y su grado de degeneración. De cualquier modo, siempre emerge una cuasiespecie dominante alrededor de la señal más fuerte.

### Población finita

Cuando un sistema genético es usado como herramienta de búsqueda siempre es necesario elegir los valores óptimos de los parámetros que se van a usar en dicho sistema como lo es la tasa de mutación  $p$  o el tamaño de la población. Se ha visto que, en el paisaje de aguja en un pajar y cuando no se tiene la aguja en la población, si  $p$  está cerca de la tasa crítica  $p^*$  de la aguja, entonces, al hallarse la aguja, no es posible formar una cuasiespecie rápidamente y hay una alta probabilidad de que la aguja se extinga en seguida lo cual es una gran pérdida ya que no es fácil volverla a encontrar. Esto indica que la  $p$  óptima debe estar ligada a la  $p^*$ , ya que no debe ser muy cercana a ella sino bastante menor, cerca de la mitad de  $p^*$ . Por su parte, se mostró que, la  $p^*$  crece logarítmicamente con la ganancia de aptitud de la aguja sobre la paja ( $G$ ) y con la robustez de la aguja, y decrece al decrecer el tamaño de la población, al aumentar el nivel de ruido en el paisaje y al aumentar la longitud del genoma. Por lo tanto, la  $p$  óptima varía de la misma manera.

Por otro lado se vio que, en el paisaje de aguja en un pajar, hay una separación total entre exploración y explotación ya que la mejor tasa para encontrar la aguja es 0.5 o búsqueda aleatoria y la mejor tasa para conservarla, una vez encontrada, es 0. La heurística estándar  $1/N$  funciona bien para buscar pero no para conservar. Se concluye entonces que una tasa óptima dependerá del estado de la población y por lo tanto es variable así que la conocida heurística  $1/N$  ya no tiene sentido, tampoco cualquier otra tasa de mutación fija.

Cuando hay dos señales, una más alta y la otra más robusta, puede aparecer una tasa  $p_f$  de transición que determina cual de las dos señales será preferida cuando ambas están en la población. Si  $p > p_f$  será la señal ancha la que domine a la población desviando al sistema lejos del genotipo óptimo, que es el que forma la señal alta. Esta  $p_f$  crece si la distancia de Hamming entre las señales es muy baja y crece de forma casi lineal con la diferencia entre la aptitud de las señales. Cuando la distancia de Hamming es alta, la presencia de la señal ancha es negativa para el proceso de búsqueda ya que hay más probabilidad de que la población sea atraída hacia ella alejándola del óptimo y requiriéndose tasas de mutación óptimas más altas que incluso pueden ser cercanas a 1. Si ambas están cerca entonces la búsqueda es mejor ya que la señal ancha coopera con la señal alta y la tasa óptima es, ahora, baja. Esta es una razón más para ver que la heurística  $1/N$  no tiene sentido porque no considera los detalles en el paisaje de aptitud.

Cuando la población encuentra primero la señal ancha (o cualquier otra señal diferente de la que contiene al óptimo) debe buscar desde allí el óptimo. Si el óptimo está suficientemente cerca, la tasa óptima es mucho menor que la tasa crítica del subóptimo, aproximadamente la mitad. Y si está lejos, la tasa óptima es cualquiera que sea mayor a la tasa crítica del subóptimo ya que la población puede así escapar de dicha señal y buscar en el resto del paisaje.

En paisajes de aptitud complejos pueden darse todas estas variantes básicas en diferentes etapas de la evolución. Entonces, es posible afirmar que ninguna tasa de mutación fija es óptima en general. Cuando se fija una tasa de mutación adecuada para un caso, se tiene necesariamente una tasa no óptima para otros casos. Sin embargo, si se ha de tener una tasa fija de mutación, en general, los GLE para determinar la tasa de mutación óptima son: la fuerza de las señales subóptimas, su cercanía con el óptimo y el estado de la población. Si en el paisaje de aptitud hay subóptimos débiles o que están cerca del óptimo se recomiendan tasas bajas de mutación y cuando hay subóptimos robustos o que están lejos del óptimo se recomiendan tasas altas. Si, en alguna generación  $t$ , la Cuasiespecie dominante está cerca del óptimo, tasas bajas son mejores y viceversa.

## Notas



Se describen el Teorema de Esquemas de Holland (HST) **¡Error! No se encuentra el origen de la referencia.** ], la Hipótesis de Bloques Constructores (BBH) **¡Error! No se encuentra el origen de la referencia.** ] y la ecuación exacta de evolución para poblaciones infinitas con recombinación tanto en la base de las secuencias genéticas como en la de los bloques constructores (que es más natural para el análisis con recombinación **¡Error! No se encuentra el origen de la referencia.** ]). Esta última ecuación contiene al HST en un contexto más general. Después se presenta el Teorema de Geiringer **¡Error! No se encuentra el origen de la referencia.** ] que describe el efecto de la recombinación sola, es decir, en un paisaje de aptitud plano y sin mutación. Se definen algunas métricas que son usadas posteriormente para aclarar el papel que juega la recombinación en los sistemas genéticos. Con estas métricas se estudia la interacción entre la recombinación y la selección bajo distintos paisajes de aptitud modelo en los que se tienen diferentes grados de modularidad que es donde se espera que la recombinación tenga efectos positivos.

Más adelante se estudia, de manera empírica y con población finita, los efectos de la recombinación con mutación y particularmente un tipo de recombinación en la que los puntos de cruce están limitados a ciertos puntos fijos del genoma y su relación con la modularidad del paisaje de aptitud.

## 2.1 Teorema de esquemas y la hipótesis de bloques constructores

La recombinación es el operador clave que distingue a los algoritmos evolutivos, como los algoritmos genéticos, de otras clases de heurísticas. La idea es que la recombinación debe tomar distintas soluciones parciales del problema y unirlos en una sola para formar una solución parcial mejorada hasta conseguir la solución total. Estas soluciones parciales se representan como esquemas. Un esquema tiene un número de bits  $l$  (o genes) definidos, es decir, que tienen un valor dado mientras que el resto de los bits están indefinidos y pueden tener cualquier valor. La longitud de un esquema  $l$  se define como el número de bits (definidos o no) que hay entre el primero y el último bits definidos incluyéndolos.

El Teorema de Esquemas de Holland (HST por sus siglas en inglés “Holland’s Schema Theorem”) **¡Error! No se encuentra el origen de la referencia.** ] establece que: para un Algoritmo Genético con cruzamiento de un punto, selección proporcional a la aptitud y mutación puntual, y para un esquema arbitrario  $I$ , con una longitud  $l$  y orden  $o$ , el valor esperado de la concentración  $P_I$  de un esquema  $I$  en la generación  $t+1$  está acotado por

$$\langle P_I(t+1) \rangle \geq p_c \frac{l+1}{N} (1-p)^o \frac{f_I(t)}{\bar{f}(t)} P_I(t)$$

donde  $\langle \rangle$  denota valor esperado,  $P_I(t)$  es la concentración del esquema  $I$  en la generación  $t$ ,  $p_c$  es la probabilidad de cruzamiento,  $p$  es la probabilidad de mutación,  $\bar{f}(t)$  es la aptitud promedio de la población en la generación  $t$  y  $f_I(t)$  es la aptitud del esquema  $I$  en la generación

$t$  dada por

$$f_I(t) \geq \frac{\sum_j f_j P_j(t)}{\sum_j P_j(t)}$$

donde las sumas son sobre todas las secuencias genéticas  $j$  que son elementos del esquema  $I$  y  $f_j$  es la aptitud de la secuencia  $j$ .

Esta versión del teorema de esquemas indica que el esquema  $I$  tiene una cota inferior de concentración en la población considerando únicamente el efecto destructivo de la recombinación y no la creación del esquema  $I$  mediante la recombinación de otros esquemas considerados sus bloques constructores.

La manera canónica de interpretar el HST es que, mientras más largo sea un esquema, menor probabilidad habrá que el esquema propague sus bits definidos en la población debido a que el operador de recombinación puede “romperlo” con un número mayor de puntos de cruzamiento posibles. Como también es más probable que los esquemas más aptos sean propagados se formaliza todo esto en la Hipótesis de Bloques Constructores (BBH por sus siglas en inglés “Building Block Hypothesis”) **¡Error! No se encuentra el origen de la referencia. ]**

? Un Algoritmo Genético funciona recombinando esquemas cortos de orden bajo y de alta aptitud hacia esquemas de mayor orden y mayor aptitud.

Esto proviene de una interpretación particular del HST en la que la ruptura de esquemas largos lleva a la idea de que sean de bajo orden. De cualquier modo, el HST no explica la yuxtaposición de estos bloques constructores en la formación de mejores esquemas.

De lo anterior surge la necesidad de una ecuación que describa de manera exacta, y no con una desigualdad, tanto la destrucción de esquemas como su construcción a partir de sus bloques constructores. A continuación se presentan dos formulaciones exactas, una en términos de secuencias genéticas y la otra, derivada de ésta, en términos de esquemas.

## 2.2 La ecuación de evolución exacta con recombinación para secuencias genéticas

Una formulación exacta de la dinámica genética bajo el operador de recombinación es útil porque da una herramienta de análisis a nivel teórico para analizar bajo qué circunstancias es bueno usar la recombinación. Una posible formulación, en términos de las secuencias genéticas, es la que a continuación se describe.

La ecuación de evolución con recombinación **¡Error! No se encuentra el origen de la referencia. ]** (sin mutación por claridad ya que la mutación se aplica por separado) en la base de las secuencias genéticas y para una población infinita es

$$P_i(t+1) = (1-p_c)P_i(t) + p_c \sum_m p_m \sum_{j,k} \varphi_i^{jk}(m) P_j(t) P_k(t) \quad (0.1)$$

donde  $P_i(t)$  es la proporción de la secuencia  $i$  al tiempo  $t$ ,  $i$  identifica a una secuencia genética, también

$$P_i(t) = \frac{f_i}{\bar{f}(t)} P_i(t)$$

es la probabilidad de seleccionar  $i$  al tiempo  $t$  (cuando ésta es proporcional a la aptitud de  $i$ ),  $f_i$  es la aptitud de  $i$ ,  $\bar{f}(t)$  es la aptitud promedio de la población al tiempo  $t$ ,  $\varphi_i^{jk}(m)$  es la probabilidad de formar  $i$  con las secuencias  $j$  y  $k$  como padres usando la máscara de recombinación  $m$  (esta probabilidad solamente puede ser 0 o 1),  $p_m$  es la probabilidad de usar la máscara  $m$  y  $p_c$  es la probabilidad de cruzamiento.

Una máscara de recombinación es una secuencia de ceros o unos de longitud igual a la del genoma en la que un 0 significa que el bit correspondiente del hijo en esa posición será tomado del primer padre y un 1 que será tomado del segundo padre.

El primer término del lado derecho de (0.1) es el término de conservación de la secuencia  $i$ , es decir, la fracción de los individuos con la secuencia  $i$  que fueron seleccionados y que no se recombinaron. El segundo es el término de reconstrucción y representa la suma de todos los individuos con la secuencia  $i$  que se formaron a partir de una recombinación de los padres  $j$  y  $k$ .

### 2.3 La ecuación de evolución exacta con recombinación para esquemas

Una formulación como la anterior resulta muy difícil de estudiar para el caso de recombinación y no da una visión clara de cuándo es útil este operador. Una manera de reformular la ecuación es ponerla en términos de esquemas en lugar de secuencias genéticas. De esta manera resulta natural el análisis del papel que juega la recombinación.

El primer paso es escribir una ecuación como la (0.1) pero para un esquema  $I$ , la cual es

$$P_I(t+1) = (1-p_c)P_I(t) + p_c \sum_m p_m \sum_{j,k} \varphi_I^{jk}(m) P_j(t) P_k(t) \quad (0.2)$$

en donde  $j$  y  $k$  son secuencias genéticas. Los términos de la suma sobre  $j$  y  $k$  pueden agruparse según la máscara que se use. Por ejemplo, para formar  $I = *1111$  ( $N = 5$ ) con la máscara  $m = 00011$ , las secuencias  $j$  que aportan (con las que  $\varphi_I^{jk} = 1$ ) pertenecen al esquema  $I_m = *11**$  y las secuencias  $k$  que aportan pertenecen al esquema  $I_{\bar{m}} = ***11$ . Entonces la ecuación puede



escribirse en términos de estos esquemas en lugar de la suma sobre las secuencias  $j$  y  $k$  reduciéndose a un solo término por cada máscara. A esto se le conoce como un *coarse graining* o granulado grueso de las ecuaciones **¡Error! No se encuentra el origen de la referencia.** **¡Error! No se encuentra el origen de la referencia.** ], quedando como

$$P_I(t+1) = (1 - p_c) P_I(t) + p_c \sum_m p_m P_{I_m}(t) P_{I_{\bar{m}}}(t)$$

Considerando que  $\sum_m p_m = 1$ , esta ecuación puede reescribirse como:

$$P_I(t+1) = P_I(t) + p_c \sum_m \rho_I(m, t) \tag{0.3}$$

donde

$$\rho_I(m, t) = P_I(t) - P_{I_m}(t) P_{I_{\bar{m}}}(t)$$

es el “coeficiente de desequilibrio epistático después de aplicar selección” (selection-weighted linkage disequilibrium coefficient **¡Error! No se encuentra el origen de la referencia.** ),  $\bar{m}$  es la máscara inversa a  $m$ ,  $I_m$  es el esquema definido con los bits de  $I$  que provienen del primer padre usando la máscara  $m$ ,  $P_{I_m}(t)$  es la proporción del esquema  $I_m$  al tiempo  $t$ ,  $p_m$  es la probabilidad de usar la máscara  $m$  y  $P'_{I_m}(t)$  es la probabilidad de seleccionar el esquema  $I_m$  al tiempo  $t$  que, si se usa selección proporcional a la aptitud, está dada por

$$P'_{I_m}(t) = \frac{f_{I_m}(t)}{\bar{f}(t)} P_{I_m}(t),$$

donde  $f_{I_m}(t) = \frac{\sum_{j \in I_m} f_j P_j(t)}{\sum_{j \in I_m} P_j(t)}$ .

La importancia de definir por separado  $\rho_I(m, t)$  es que ésta determina si, para cada máscara  $m$ , la recombinación será más constructiva (cuando sea menor que 0) o más destructiva (cuando sea mayor que 0) en función de la distribución de la población en la generación  $t$ . Esto debe ser claro al observar (0.3).

Ejemplo: Con  $N = 6$ , para el esquema **\*\*0010**, si se usa cruzamiento en un punto con igual probabilidad ( $p_m = 1/5$ ) para cada punto de cruce la ecuación es

$$\begin{aligned}
& P_{**0010}(t ? 1) ? P_{**0010}^?(t) ? p_c \frac{1}{5} [ \\
& \quad ? P_{**0010}^?(t) ? P_{**0**}^?(t) P_{**010}^?(t) ? \\
& \quad ? ? P_{**0010}^?(t) ? P_{**00**}^?(t) P_{**10}^?(t) ? \\
& \quad ? ? P_{**0010}^?(t) ? P_{**001*}^?(t) P_{**0}^?(t) ? \\
& ]
\end{aligned}$$

A su vez, la ecuación para uno de los esquemas que intervienen en esta ecuación se forma de la misma manera, por ejemplo, la ecuación para  $**010$  sería

$$\begin{aligned}
& P_{**010}(t ? 1) ? P_{**010}^?(t) ? p_c \frac{1}{5} [ \\
& \quad ? P_{**010}^?(t) ? P_{**0**}^?(t) P_{**10}^?(t) ? \\
& \quad ? ? P_{**010}^?(t) ? P_{**01*}^?(t) P_{**0}^?(t) ? \\
& ]
\end{aligned}$$

y así sucesivamente hasta llegar al nivel mas primitivo que son los esquemas de orden 1 los cuales no pueden ser formados con esquemas de menor orden y sus ecuaciones son de la forma

$$P_{**0**}(t ? 1) ? P_{**0**}^?(t)$$

que es, naturalmente, independiente de  $p_c$  ya que estos esquemas son los objetos primitivos de la recombinación.

A estos esquemas particulares que intervienen en la ecuación de un esquema  $I$  dado se les llama Bloques Constructores de  $I$ .

Dado un esquema  $I$  de orden  $N$ , es decir, una secuencia genética, y un operador de recombinación se puede escribir la ecuación que describe su dinámica en función de sus Bloques Constructores únicamente y, a su vez, las ecuaciones de estos Bloques Constructores en función de sus Bloques Constructores, etc. El conjunto de esquemas que intervienen en estas ecuaciones, derivadas a partir de  $I$ , forman una base con la que puede deducirse la dinámica de cualquier otra secuencia o esquema. A esta base se le denomina la Base de Bloques Constructores (BBB por sus siglas en inglés “Building Block Basis”). Para cada secuencia genética existe una BBB asociada a ésta y cada una de estas bases son equivalentes entre sí. Una base muy importante es la que está asociada a la secuencia genética del óptimo cuyos componentes son los únicos que intervienen en la dinámica para encontrarla.

## 2.4 Métricas

Las métricas que a continuación se describen serán usadas en las siguientes secciones en el análisis del efecto de la recombinación.

### 2.4.1 Distancia al centro del simplex

Se denomina el *simplex* al conjunto de soluciones de

$$\sum_{i=1}^{2^N} P_i(t) = 1, \quad P_i \geq 0$$

donde cada solución es un punto de un hiperplano de dimensión  $2^N - 1$  que está dentro de un espacio vectorial de dimensión  $2^N$ . Las coordenadas de cada punto solución son las  $P_i(t)$  que, en este trabajo, representan a las concentraciones de cada secuencia genética  $i$  en una población. Para toda  $t$ , las concentraciones de las secuencias cumplen esta ecuación y por lo tanto son puntos en el simplex.

El *centro del simplex* es el caso cuando la población está uniformemente distribuida en el dominio de los genotipos, es decir, cuando todos los genotipos ocupan la misma proporción de la población que es  $P_i = 1/2^N \forall i$ . Se define la distancia al centro del simplex como la distancia entre este punto y el estado actual de la población en la generación  $t$  y se calcula como

$$dC(t) = \sqrt{\sum_{i=1}^{2^N} \left( P_i(t) - \frac{1}{2^N} \right)^2}.$$

### 2.4.2 Distancia a la variedad de Geiringer

La variedad de Geiringer es el conjunto de puntos de equilibrio posibles en un sistema con población infinita que evoluciona únicamente mediante recombinación. Cada población inicial se asocia con uno, y solo un, punto en la variedad de Geiringer pero no a la inversa ya que varias poblaciones iniciales se asocian con el mismo punto en la variedad de Geiringer. Se define  $G_i(t)$  como la proporción de secuencias genéticas  $i$  que se tendrán en la variedad de Geiringer a partir del estado de la población en la generación  $t$ . Esta cantidad está dada por el teorema de Geiringer (ver la sección 2.6.1 más adelante).

Se define ahora una métrica que sirve para medir qué tan lejos está la población de la variedad de Geiringer como

$$dG(t) = \sqrt{\sum_{i=1}^{2^N} \left( P_i(t) - G_i(t) \right)^2}$$

donde  $G_i$  es la proporción de la población que ocuparía la secuencia  $i$  en la variedad de Geiringer a partir de la población actual.

### 2.4.3 *Distancia al óptimo*

Esta métrica mide la distancia a la que se encuentra la población de estar formada únicamente por copias del genotipo del óptimo  $i_0$  y está dada por

$$dO(t) = \sqrt{\sum_{i \neq i_0} P_i(t)^2}.$$

En caso de ser un paisaje plano,  $i_0$  no es el óptimo sino una secuencia de referencia solamente.

### 2.4.4 *Frecuencia relativa por efecto de la recombinación*

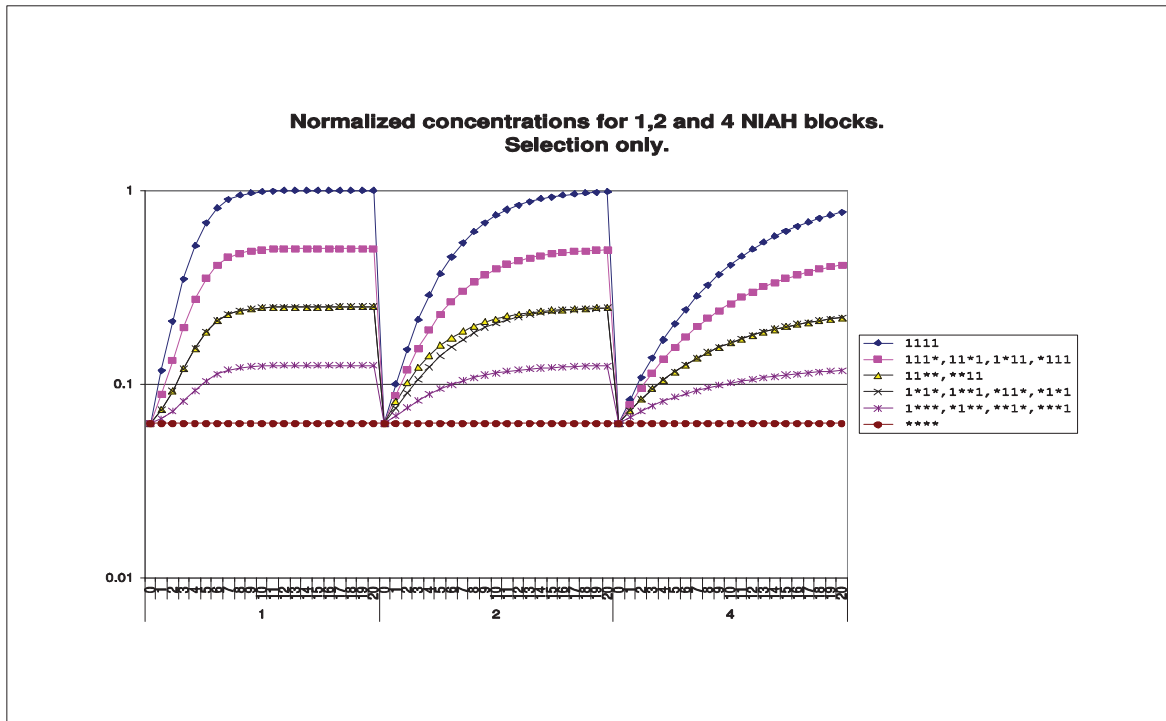
Esta métrica compara la concentración del esquema  $I$  al evolucionar con selección y recombinación ( $s+r$ ) contra la concentración al evolucionar sin recombinación. Se define como

$$\frac{P_I^{(s+r)}(t)}{P_I^{(s)}(t)}.$$

## 2.5 Paisajes de aptitud modulares

Un paisaje de aptitud modular consiste en la agrupación de los genes en  $M$  bloques en donde cada bloque  $\beta$  tiene un número de genes sobre los cuales está definido un paisaje de aptitud. La aptitud total  $f_i$  de un genotipo  $i$  es la suma de sus aptitudes en cada uno de los bloques.

$$f_i = \sum_{\beta=1}^M f_{i,\beta} \tag{0.4}$$



**Figura 2-1** Concentraciones normalizadas (por el número de secuencias en cada esquema) de los esquemas que son potencialmente bloques constructores de la secuencia óptima 1111 en función del tiempo aplicando únicamente selección en paisajes con 1, 2 y 4 bloques de aguja en un pajar de 4,2 y 1 bit respectivamente y a partir de una población infinita uniforme.

En estos casos el óptimo es obviamente el genotipo que contiene al óptimo de cada bloque. Para el caso cuando los bloques son de aguja en un pajar, el óptimo sería el genotipo con cada una de las agujas. Cuando los bloques están formados únicamente por genes contiguos en el genoma se dice que el paisaje es una concatenación de bloques.

En la **Figura 2-1** se muestra la evolución (a partir de una población infinita uniformemente distribuida sobre el espacio de los genotipos) de los esquemas que se derivan de la secuencia óptima 1111 ( $N = 4$ ) en paisajes con 1,2 y 4 bloques concatenados de aguja en un pajar cada uno con 4,2 y 1 gen respectivamente sin aplicar mutación ni recombinación. En todos los casos se nota una preferencia por los esquemas de mayor orden. Puede verse que para el caso con 2 bloques hay una preferencia por 11\*\* sobre 1\*1\*, es decir, por aquellos esquemas que coinciden con los bloques definidos en el paisaje a pesar de ser del mismo orden. Obviamente, en los casos extremos en donde hay un solo bloque y también en donde cada gen es un bloque no puede darse esto ya que tanto 11\*\* como 1\*1\* tienen en ambos casos el mismo número de óptimos de bloque. En otras palabras, no solo se prefiere a los esquemas que tengan más coincidencias con los genes del óptimo sino también a los que coincidan con el óptimo de cada bloque.

Los paisajes de aptitud modulares son frecuentes en la naturaleza. Por ejemplo, un rasgo fenotípico como la forma de una oreja en alguna especie, se codifica en alguna región del

genoma específica. Si esta región cambia su configuración genética, también cambiará el rasgo fenotípico asociado sin afectar a otros. Así, esta región del genoma puede evolucionar independientemente de las demás hasta hallarse la configuración óptima. Al optimizarse cada módulo independientemente se llega a un óptimo global simplemente por yuxtaposición de los módulos optimizados.

## 2.6 Efecto de la recombinación sola

Para estudiar la recombinación es importante entender cuál o cuáles son los sesgos intrínsecos del operador por sí mismo para luego poder distinguir sus efectos cuando se aplica junto con otros operadores.

Se describe ahora el Teorema de Geiringer que indica cual es el punto de equilibrio del sistema bajo el efecto de la recombinación sola. En seguida se hace un estudio a través del tiempo de este efecto sobre un paisaje de aptitud plano sin mutación mediante un conjunto de métricas para evaluarlo y a partir de diferentes distribuciones de población inicial extremas.

### 2.6.1 Teorema de Geiringer

El teorema de Geiringer **¡Error! No se encuentra el origen de la referencia.** ] establece que, si se aplica solamente el operador de recombinación sin selección (paisaje de aptitud plano) y sin mutación, entonces la probabilidad de que un esquema  $I$  sea seleccionado en estado estacionario está dada por

$$P_I(t \rightarrow \infty) = \prod_{r=1}^N P_{I_{*r}}(0), \quad I_{*r} = \dots * I_r * \dots$$

donde  $I$  es un esquema e  $I_{*r}$  es el esquema de orden 1 definido solamente con el  $r$ -ésimo bit  $I_r$  de  $I$ . Es decir, que la proporción del esquema  $I$  tiende a un valor que depende únicamente de las proporciones iniciales de cada esquema de orden uno definido mediante los bits de  $I$ . Para cada distribución inicial de la población existe entonces una distribución en el punto de equilibrio única. Al conjunto de distribuciones en el punto de equilibrio posibles se le llama la “variedad de Geiringer”.

### 2.6.2 Resultados

Ahora se analizan diferentes casos de inicialización de la población con  $N = 6$  para ver el efecto que tiene la recombinación en cada uno sobre un paisaje de aptitud plano (selección nula). Estos casos son:

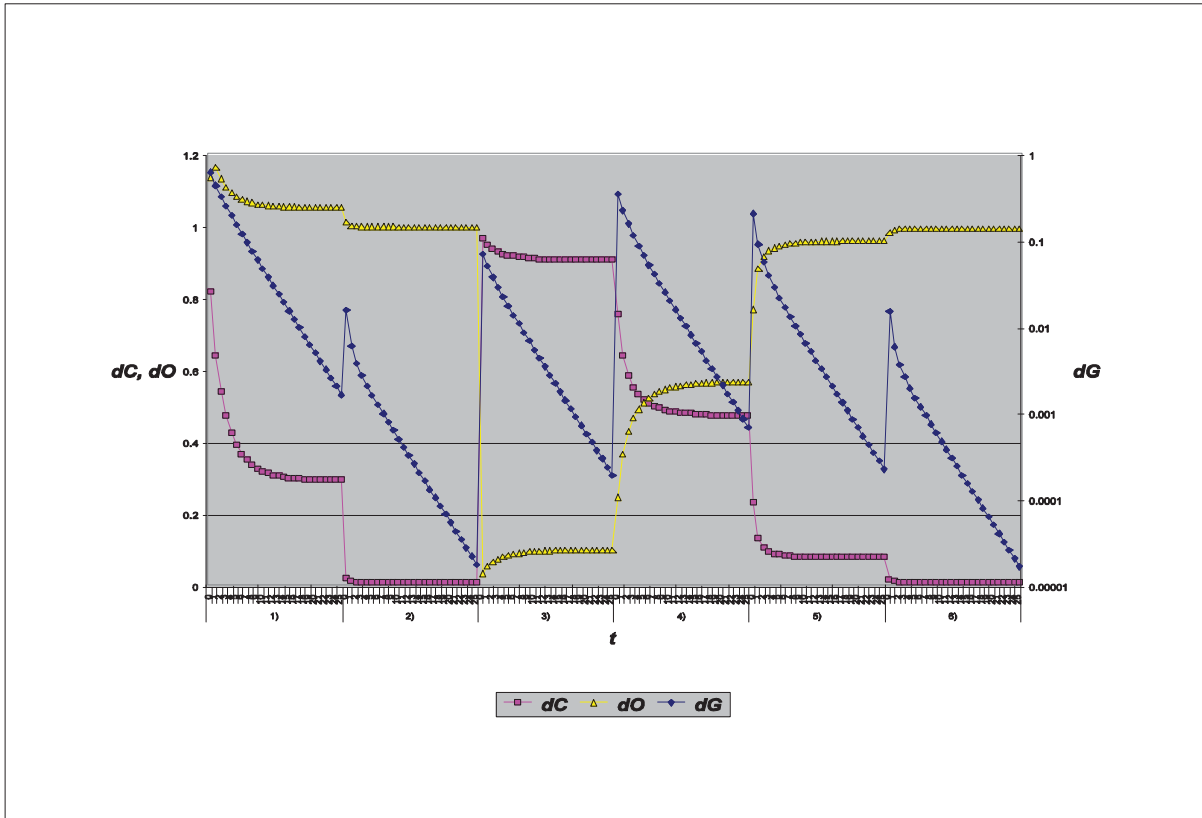
- 1)  $P_{\text{aguja}} = 0.2$  y  $P_{\text{opuesta}} = 0.8$ ,  $P_{\text{otra}} = 0$ . Opuestos distintos.
- 2)  $P_{\text{aguja}} = 0$  y  $P_{\text{otra}} = 1/(2^N - 1)$ . Distribución uniforme pero sin la aguja.

- 3)  $P_{\text{aguja}} = 2000P_{\text{otra}}$ . Dominan mucho las agujas (96% de la población).
- 4)  $P_{\text{aguja}} = 200P_{\text{otra}}$ . Agujas dominan (75%) pero hay bastante paja.
- 5)  $P_{\text{aguja}} = 20P_{\text{otra}}$ . Las agujas dominan poco (23%).
- 6)  $P_{\text{aguja}} = 2P_{\text{otra}}$ . Las agujas casi no dominan (3%).

En todos los casos se observa un acercamiento exponencial a la variedad de Geiringer que, después de un transitorio de pocas generaciones, este acercamiento lleva la misma intensidad. Véase en la **Figura 2-2**  $dG$  en la escala logarítmica de la derecha.

Cuando la población es muy homogénea (dominada por algún genotipo) o cuando es aleatoria (o casi aleatoria), el efecto de la recombinación es pequeño en cuanto a los cambios en la distribución de la población sobre el espacio de los genomas, es decir, la población permanece dominada por el mismo genotipo o permanece aleatoria. Véase la distancia al centro del simplex (que es la población aleatoria)  $dC$  en la **Figura 2-2** en los casos 2), 5) y 6) en donde la población es bastante aleatoria y el caso 3) en donde la población está muy dominada por el óptimo.

Cuando la población está dominada por algún genotipo pero no totalmente de manera que una buena parte de la población mantiene una alta diversidad genética, entonces el efecto de la recombinación es grande y con tendencia hacia el centro del simplex, aunque no lleva hasta allá totalmente. Véase la **Figura 2-2** en los casos 1) y 4). En estos casos, al recombinarse, la secuencia dominante “contagia” a los demás individuos de sus genes disminuyéndose su dominio. Mientras más contagio haya menor será el efecto de la recombinación ya que la destrucción de la secuencia dominante será compensada con la construcción de ésta mediante recombinaciones de los individuos contagiados de sus genes. Ambos procesos se igualan cuando la población alcanza la variedad de Geiringer.



**Figura 2-2** Métricas  $dG$ ,  $dC$  y  $dO$  en función del tiempo aplicando recombinación en un paisaje plano y sin mutación para los distintos casos de inicialización y  $N = 6$ .

Se pueden entonces distinguir tres distintos efectos de la recombinación sola sobre una población.

1. Efecto nulo o casi nulo cuando la población es muy dominada por algún genotipo.
2. Efecto máximo de transmisión de genes desde el genoma dominante hacia el resto de la población cuando domina un genotipo pero no totalmente.
3. Efecto de búsqueda aleatoria cuando la población es aleatoria (o casi aleatoria), es decir, solamente produce cambios aleatorios en la población.

Se identifica aquí un GLE que es el grado de dominio de un genotipo sobre la población en combinación con el grado de diversidad genética en la misma.

## 2.7 Efecto de la recombinación con mutación

En la **Figura 2-3** se muestra el análogo de la **Figura 2-2** pero para el caso de inicialización i) solamente, y se comparan los casos sin y con mutación. Cuando no se aplica ni recombinación ni mutación se tiene que todas las distancias permanecen constantes. Al aplicar recombinación se ve que la distancia a la variedad de Geiringer se reduce exponencialmente y  $dG \sim 0.67^t$ . En



contraste, la distancia al centro del simplex inicialmente se reduce pero se estabiliza en cierto valor asintóticamente. Con mutación sola tanto  $dG$  como  $dC$  se acercan a cero exponencialmente, con  $dG \sim 0.64^t$  y  $dC \sim 0.80^t$  (para la tasa de mutación usada de 0.1). Este acercamiento es más lento con tasas de mutación más bajas y viceversa. Al aplicar ambos operadores se observa una superposición de ambos efectos resultando en un acercamiento aun más rápido hacia la variedad de Geiringer, no así, al centro del simplex. Nótese que la mutación sola puede llevar a la variedad de Geiringer, pero la recombinación sola no lleva al centro del simplex. Sin embargo, la mutación siempre lleva al mismo punto en la variedad de Geiringer que es en el centro del simplex. Esto implica que, en poblaciones finitas, la recombinación llevará a cabo cambios en la genética de la población localizados alrededor de la variedad de Geiringer mientras que la mutación lo hará alrededor del centro del simplex. Lo más importante aquí es que la recombinación no tiene ningún efecto en contra de la mutación. En otras palabras, la acción de la mutación tiene una componente en la misma dirección que la recombinación, hacia la variedad de Geiringer más, aparte, otra componente en dirección al centro del simplex. Esto es importante ya que se creía que la recombinación tenía un efecto concentrador lo cual no es posible si no se opone al efecto de la mutación. Entonces, al aplicar mutación, la única manera de contrarrestar la dispersión de los individuos en el espacio de los genotipos es por efecto de la selección. Sin selección, la mutación termina por llevar a la población siempre al centro del simplex. El GLE para determinar el grado de concentración o dispersión en la población es el balance entre selección y mutación, sin que la recombinación tenga influencia en él.

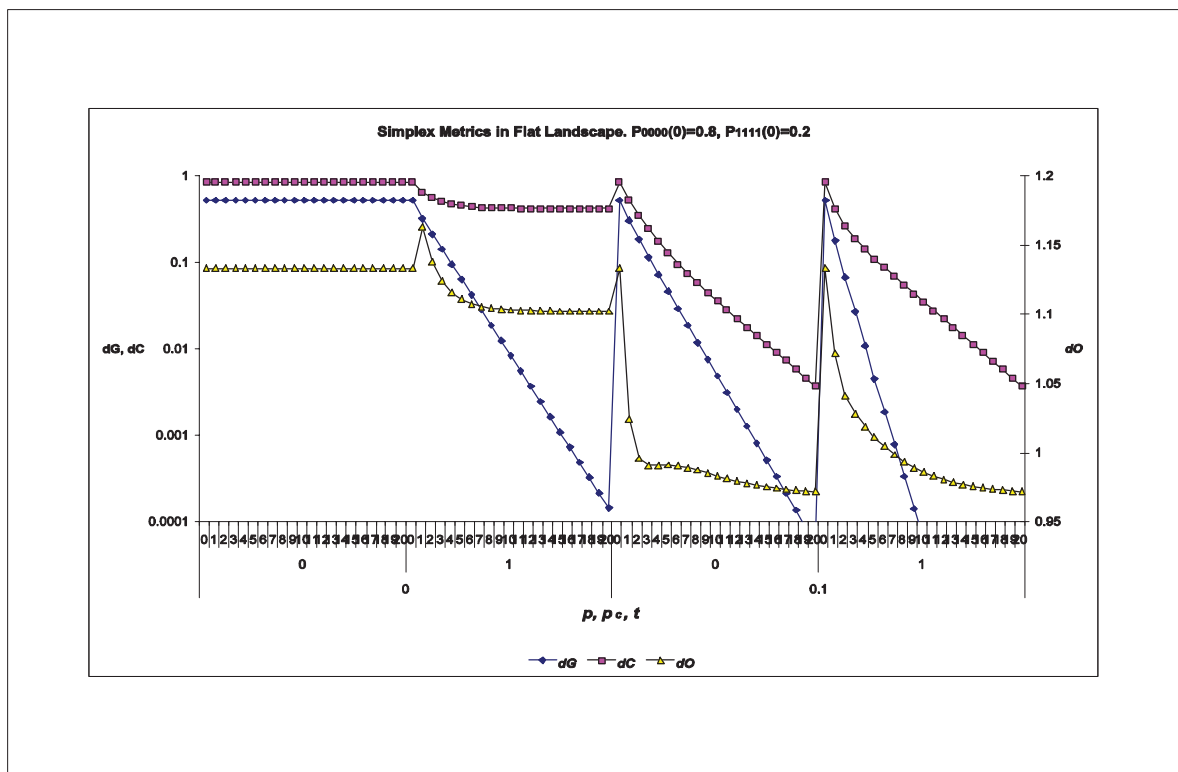
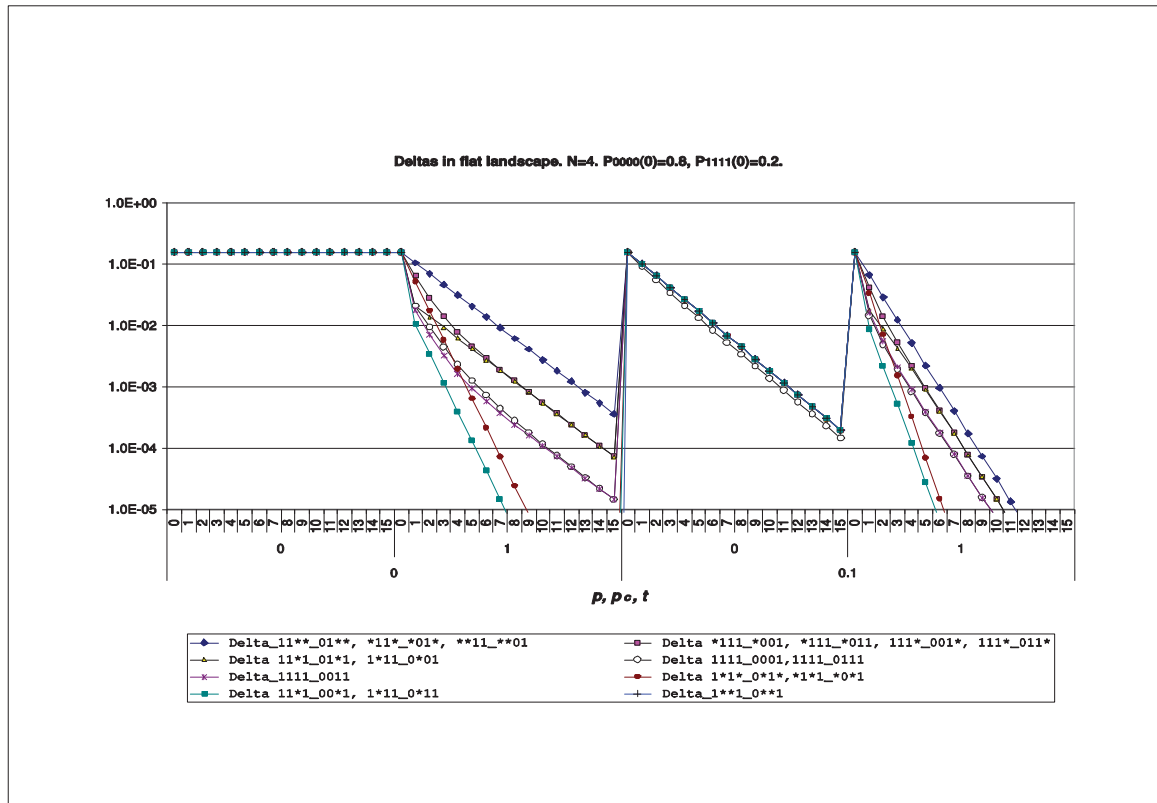


Figura 2-3 Métricas  $dG$ ,  $dC$  y  $dO$  sin/con recombinación/mutación en un paisaje de aptitud plano  $N = 4$  para el caso de inicialización 1) Opuestos distintos.



**Figura 2-4** Gráfica de  $\Delta_I(m,t)$  en un paisaje plano en el caso de inicialización i) sin/con recombinación/mutación. La notación indica el esquema  $I$  en los primeros 4 símbolos y la máscara  $m$  en los siguientes 4 símbolos.  $N = 4$ .

En la **Figura 2-4** se muestra el valor de  $\Delta_I(m,t)$  para los mismos casos de la **Figura 2-3** considerando esquemas definidos a partir de la secuencia 1111 y máscaras de recombinación con cruzamiento en un punto.

Sin operadores, todos los valores se mantienen constantes. Con recombinación sola, todas las curvas decrecen exponencialmente ya que la población se va aproximando a la variedad de Geiringer en donde por definición  $\Delta_I(m,t)$  es 0, es decir, que, en ese punto, la recombinación construye y destruye el mismo número de esquemas  $I$  lo cual significa que está en equilibrio y la recombinación ya no tiene efecto neto. La única excepción aquí es para esquemas de orden uno, en donde  $\Delta_I(m,t)$  es siempre cero ya que no puede haber destrucción ni reconstrucción por recombinación.

Las curvas para 11\*\*, 111\* y 1111 decaen a la misma tasa (después de un corto transitorio) para cualquier máscara. Por otro lado la curva para 1\*1\* decae más rápidamente ya que se tienen 2 máscaras, y no una, con las cuales es posible destruir el esquema. La tasa de decaimiento para el esquema 11\*1 depende de la máscara usada. Si la máscara corta el par de 1s contiguos, en este caso la máscara 0111, la tasa de decaimiento es igual a la del esquema 11\*\* mientras que, si la máscara es una de 00\*1, ésta corta el esquema entre los 1s que están separados y la tasa de decaimiento iguala a la del esquema 1\*1\*. El GLE para medir cuanto

efecto destructivo tiene la recombinación sobre un esquema dado es entonces el número de máscaras que dividen al esquema en los mismos subesquemas o bloques constructores.

Para el caso con mutación sola se observa que las mediciones tienden a cero exponencialmente a la misma tasa. Con ambos operadores se observa una superposición de sus efectos resultando en tasas de acercamiento a cero aún mayores. Esto significa que la mutación acelera el efecto de la recombinación al llevar a la población más rápidamente hacia la variedad de Geiringer pero al mismo tiempo también la acerca al centro del simplex.

## 2.8 Efecto de la recombinación con selección

Habiendo visto ya cual es el sesgo intrínseco que el operador de recombinación tiene sobre la distribución de la población y la manera en que se ve afectada por la mutación, el siguiente paso es, naturalmente, el análisis de este sesgo pero bajo diferentes paisajes de aptitud, es decir, actuando al mismo tiempo que la selección y sin mutación.

Para este análisis se usan paisajes de aptitud con  $N = 4$  de 1, 2 y 4 bloques de aguja en un pajar concatenados de 4, 2 y 1 bit cada bloque respectivamente.

### 2.8.1 Paisaje de 1 bloque de Aguja en un pajar

Con este paisaje e iniciando con una población infinita aleatoria (o uniformemente distribuida en el espacio de los genotipos), la selección hará que la aguja se reproduzca más rápido que las demás secuencias, lo que provoca que la recombinación tenga algún efecto no aleatorio sobre la población, como se vio anteriormente. Este efecto se muestra en la **Figura 2-5** en donde es notable que, para todos los esquemas derivados del genotipo óptimo (1111) o aguja, el efecto de la recombinación es el de reducir la frecuencia de éstos en la población (temporalmente). El esquema que sufre mayores pérdidas es el de mayor orden (1111) y los que menos los de menor orden (por ejemplo 1\*\*\*). Considerando esquemas del mismo orden, los de mayor longitud pierden más, por ejemplo, 11\*\* es preferido sobre 1\*1\* ya que el segundo tiene 2 puntos de cruzamiento posibles para destruirlo.

Aunque eventualmente, con o sin recombinación, la población es dominada por el óptimo, este dominio es más lento con recombinación. Se puede decir que la recombinación no ayuda a la evolución en el sentido de que el número de secuencias óptimas es menor al usarla (cuando el óptimo ya está presente) pero ayuda en el sentido de que “contagia” a la población con la genética del óptimo. Esta reducción en la frecuencia del óptimo no cambia de sentido la presión selectiva ya que, debido al contagio de sus genes en la población y a que el paisaje es plano más allá del óptimo, este contagio permanece intacto y no se extingue. Así, la genética del óptimo se queda siempre con una ventaja sobre la de las demás secuencias y termina finalmente dominando en la población.

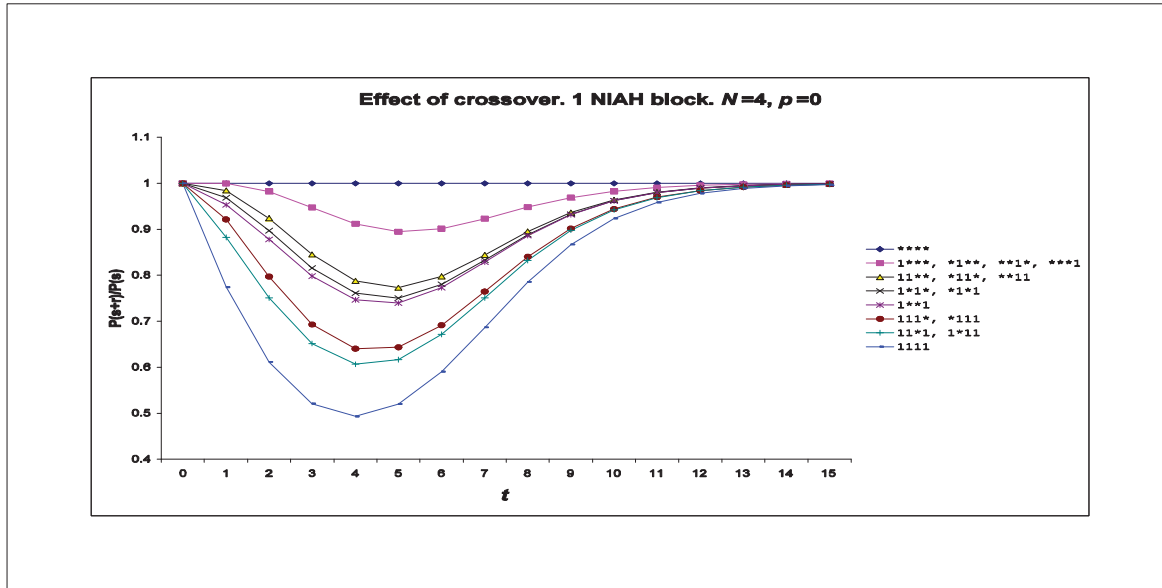


Figura 2-5 Gráfica de la frecuencia relativa (ver 2.4.4 en la página 10) por efecto de la recombinación para los esquemas derivados del óptimo en un paisaje de aguja en un pajar.

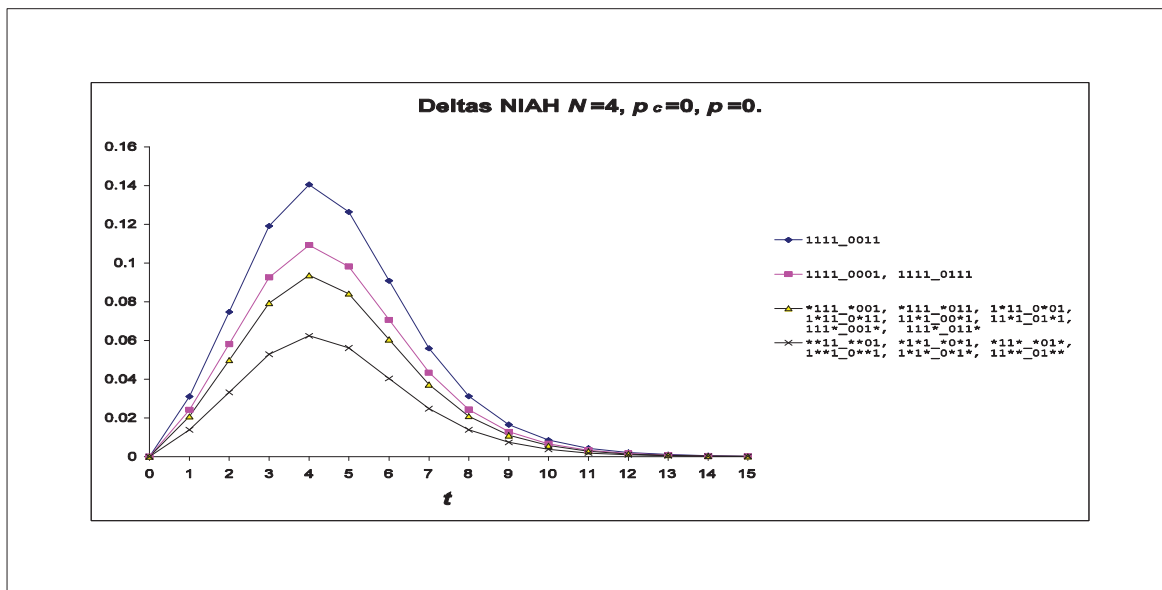
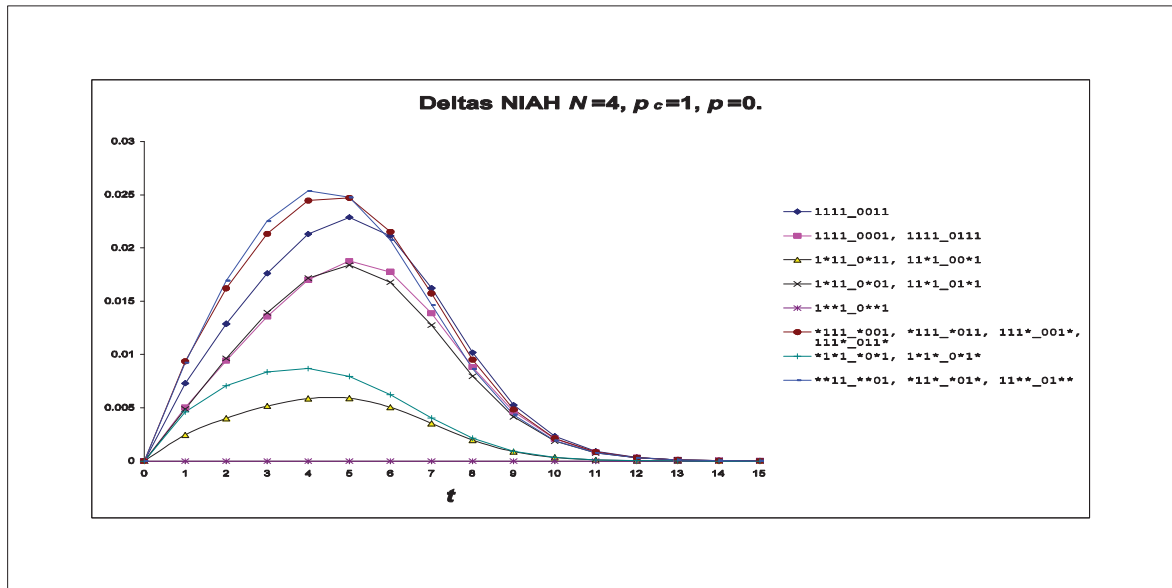


Figura 2-6 Gráfica de  $\delta_I(m,t)$  en el paisaje de aguja en un pajar  $N = 4$  sin recombinación ni mutación con población inicial infinita aleatoria. La notación es  $I_m$ , es decir, el esquema y la máscara de recombinación.

En la Figura 2-6 se muestra el valor de  $\delta_I(m,t)$  para los mismos esquemas de la Figura 2-5 en el caso en donde no se aplica recombinación midiéndose así cuánto efecto (destrutivo) tendría la recombinación de aplicarse en la generación  $t$ . Se puede ver que en todos los casos los valores son siempre positivos lo que indica que, de aplicarse recombinación, ésta será más

destruictiva que constructiva produciéndose una reducción en la frecuencia de cada esquema  $I$ .

En la **Figura 2-7** se muestran las  $\Delta_I(m,t)$  pero ahora usando recombinación. Se observa un comportamiento similar al de la **Figura 2-6** pero con valores menores. Esto se debe a que la recombinación esparce por toda la población la genética del óptimo siempre y cuando se tenga una mayor presencia de éste en la población. Este esparcimiento le cuesta al óptimo parte de esa ventaja que tenía y, entonces, la siguiente aplicación de la recombinación tiene un efecto menos destructivo. Ya sea con o sin recombinación, a la postre, todas las curvas se estabilizan en 0 indicando que la población ha sido dominada totalmente por el óptimo.

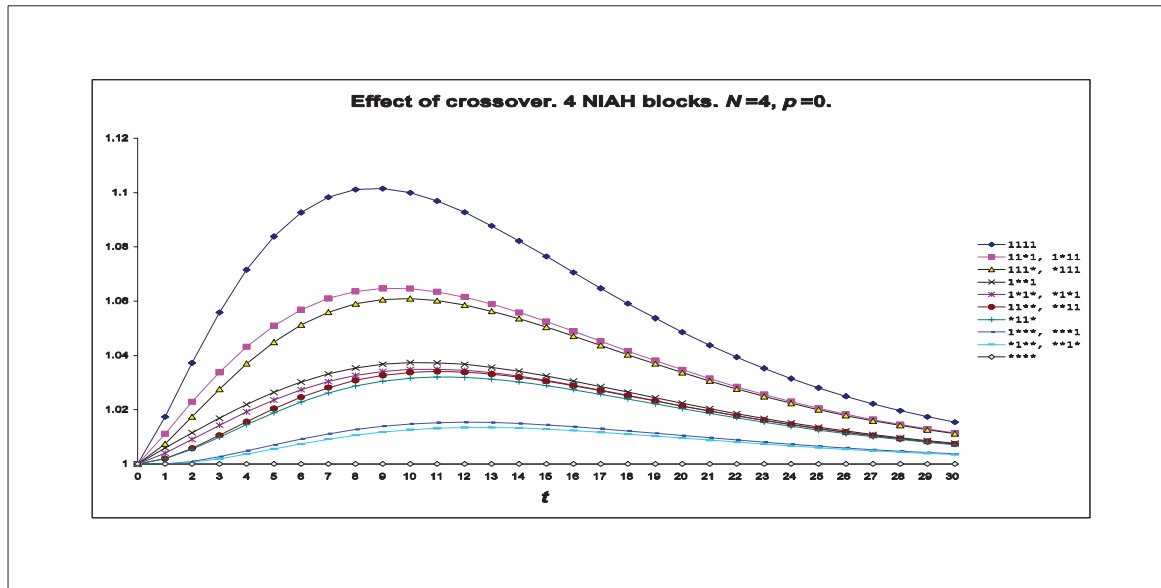


**Figura 2-7** Gráfica de  $\Delta_I(m,t)$  en el paisaje de agujas en un pajar  $N = 4$  con recombinación sin mutación con población inicial infinita aleatoria. La notación es  $I_m$ , es decir, el esquema y la máscara de recombinación.

En general, cuando se aplica recombinación, el efecto es el de reducir la ventaja de la secuencia dominante (si no domina mucho) al esparcir su genética en la población. Luego, la selección tiende a restaurar esta ventaja por lo que el efecto de la recombinación sigue siendo destructivo. Cuando toda la población queda contagiada de la genética del óptimo, la recombinación empieza a ser menos destructiva porque las recombinaciones se hacen entre individuos cada vez más similares. Con selección fuerte, la recombinación deja de tener efecto pronto y con selección muy débil el efecto es notable durante más tiempo.

### 2.8.2 Paisaje de $N$ bloques de Aguja en un pajar

En este paisaje cada bit es un bloque de agujas en un pajar. El óptimo global es el genotipo en el que se tiene la agujas (el bit óptimo) de cada bloque.



**Figura 2-8** Gráfica de la frecuencia relativa por efecto de la recombinación para los esquemas derivados del óptimo en un paisaje de 4 bloques de aguja en un pajar con  $N = 4$ .

En este caso la recombinación tiene un efecto que favorece la presencia del óptimo más que su destrucción (Véase la **Figura 2-8**). Esto es porque, cuando un individuo obtiene una ventaja reproductiva por coincidir con el óptimo de uno de los bloques, su genética es esparcida bit por bit entre la población mediante la recombinación dándole a todos los individuos, contagiados de ese óptimo de bloque, la misma ventaja. Pero, al mismo tiempo, otros individuos, que coinciden con otro óptimo de bloque, esparcen también su genética. Al esparcirse ambos óptimos de bloque, llegará un momento en que haya individuos con ambos óptimos de bloque obteniendo una mayor ventaja que aquellos que solo tienen un óptimo y, a su vez, éstos esparcirán su genética también generándose más individuos con ésta doble ventaja. De esta manera se construye el óptimo global cuando cada óptimo de bloque ha sido esparcido entre la población.

Este paisaje hace que la recombinación en un punto favorezca a los esquemas de mayor orden y a los esquemas en los que la distancia promedio de sus bits definidos con el extremo más cercano del genoma sea menor. Por ejemplo,  $1***$  es más favorecido que  $*1**$  por tener el 1 en el extremo del genoma, el esquema  $1**1$  es más favorecido que  $11**$  porque sus dos bits definidos están en el extremo del genoma,  $11*1$  es preferido sobre  $111*$  porque tiene un solo bit lejos de los extremos. Este efecto es claro en la **Figura 2-8** y se debe al tipo de recombinación que se usó que es de un punto. Con recombinación anular de dos puntos este efecto no es visible quedando solamente una preferencia por los esquemas de mayor orden sobre los de menor orden. La diferencia se explica porque los bits definidos en esquemas de orden 1 pueden ser separados de los demás con mayor a menor número de recombinaciones. Por ejemplo, en el esquema  $1***$  el 1 puede ser separado de los demás con una sola recombinación mientras que en  $*1**$  se deben usar al menos 2 recombinaciones. Esto se traduce en una mayor o menor facilidad para esparcir esos bits en la población y de recombinarse con otros bit buenos (óptimos).

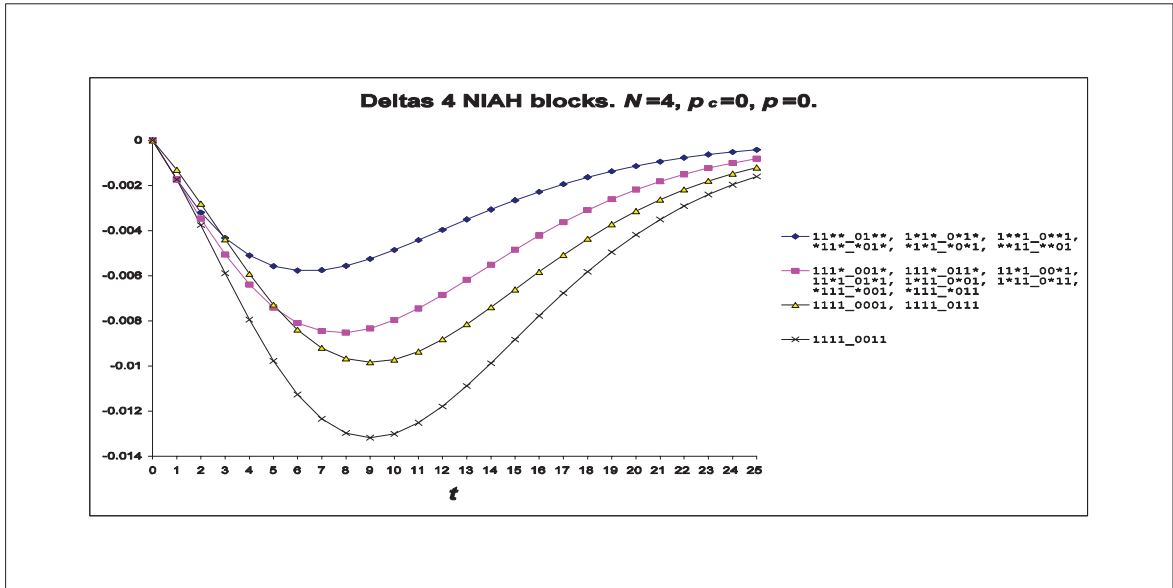


Figura 2-9 Gráfica de  $\Delta_I(m, t)$  en el paisaje de 4 bloques de aguja en un pajar  $N = 4$  sin recombinación ni mutación con población inicial infinita aleatoria. La notación es  $I_m$ , es decir, el esquema y la máscara de recombinación

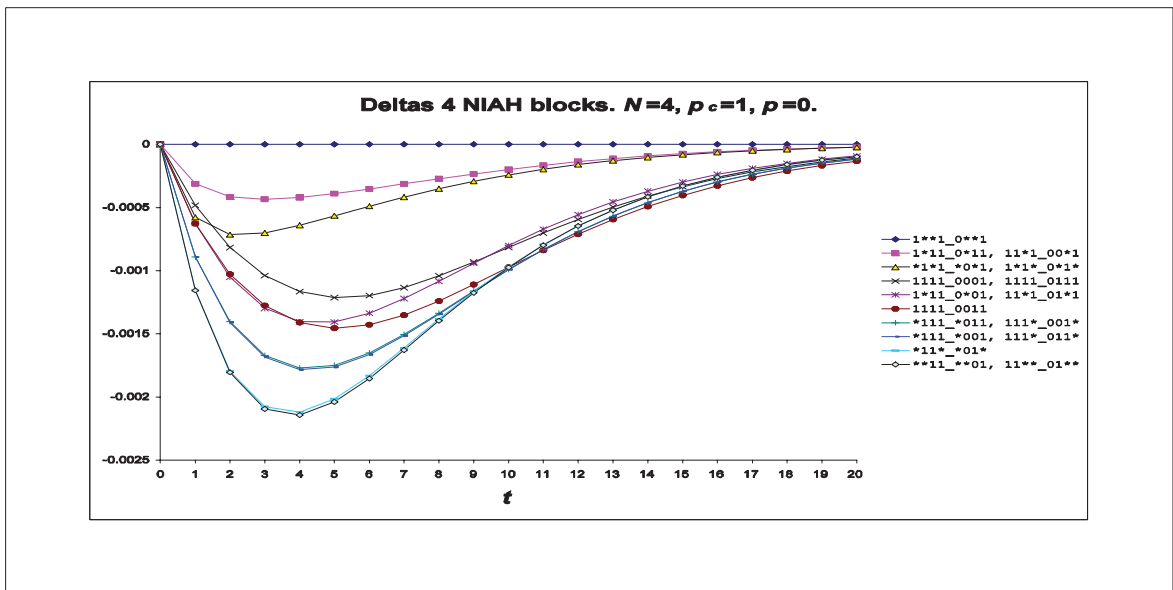


Figura 2-10 Gráfica de  $\Delta_I(m, t)$  en el paisaje de 4 bloques de aguja en un pajar  $N = 4$  con recombinación sin mutación con población inicial infinita aleatoria. La notación es  $I_m$ , es decir, el esquema y la máscara de recombinación.

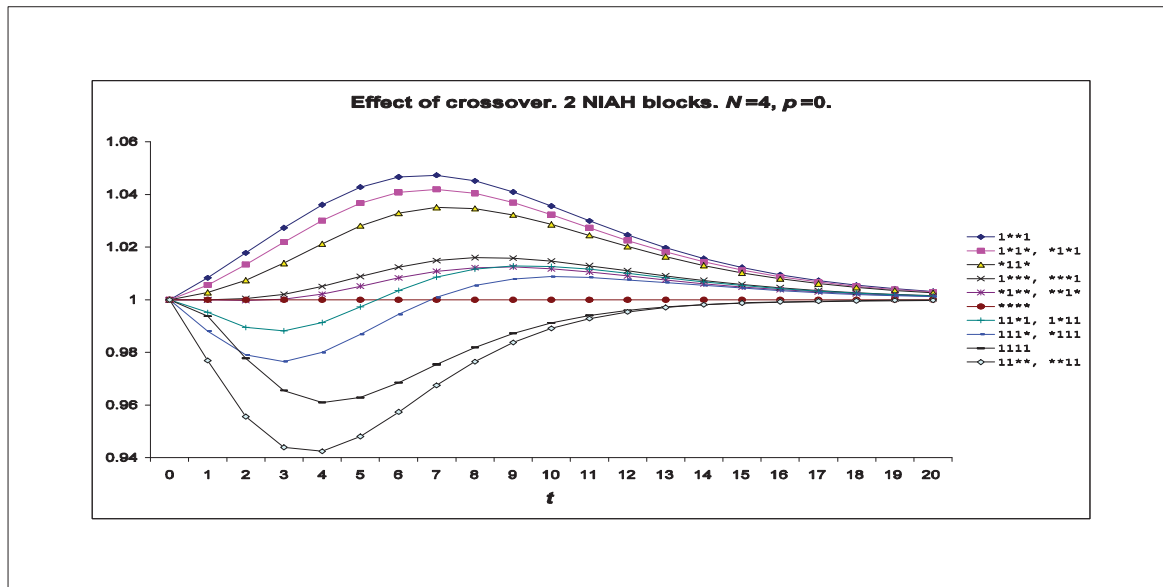
Este efecto no es visible en el paisaje anterior de un solo bloque de aguja en un pajar porque, en ese caso, los bits individuales no proporcionan ninguna ventaja siendo necesario que la combinación óptima de bits esté presente para tener alguna ventaja.

Por otro lado, en cuanto a las  $\varphi_I(m,t)$ , puede verse el efecto contrario que en el paisaje de aguja en un pajar. En la **Figura 2-9** se refleja esto en los valores siempre negativos de  $\varphi_I(m,t)$  para toda  $I$  y en toda  $t$  lo que significa que, de aplicarse la recombinación, habría mayor concentración de cada esquema del óptimo.

Cuando se aplica recombinación, las  $\varphi_I(m,t)$  siguen teniendo un comportamiento similar aunque con valores absolutos menores (**Figura 2-10**). La razón de esto es similar al caso de aguja en un pajar. Si la recombinación va a tener un efecto más constructivo que destructivo de un esquema  $I$  dado, se debe a que la fracción que sería destruida de  $I$  es menor que la fracción de la “materia prima” para construirlo que efectivamente se recombinaría y lo construiría. Al aplicar la recombinación, esta materia prima se consume en crear el esquema  $I$  y, en la siguiente generación, ya no habrá tanta de esta construcción.

### 2.8.3 Dos bloques concatenados de Aguja en un pajar

Aquí el paisaje consiste en la concatenación de 2 bloques de aguja en un pajar cada uno de 2 bits. Como puede observarse en la **Figura 2-11**, la recombinación tiene tanto efectos positivos para algunos esquemas como efectos negativos para otros. Lo mismo puede observarse en la **Figura 2-12** y en la **Figura 2-13**.



**Figura 2-11** Gráfica de la frecuencia relativa por efecto de la recombinación para los esquemas derivados del óptimo en un paisaje de 2 bloques de aguja en un pajar con  $N = 4$ .

El beneficio de la recombinación ahora depende de la pareja esquema-máscara ( $I-m$ ) y hay máscaras buenas y malas para un mismo esquema. Por ejemplo: para el esquema 11\*1 la máscara 0111 es mala ya que destruye el óptimo del primer bloque, mientras que para el mismo esquema la máscara 00\*1 es benéfica ya que respeta los bordes de los bloques. Las



máscaras que respetan la modularidad del paisaje son benéficas y las que cortan en medio de algún bloque no.

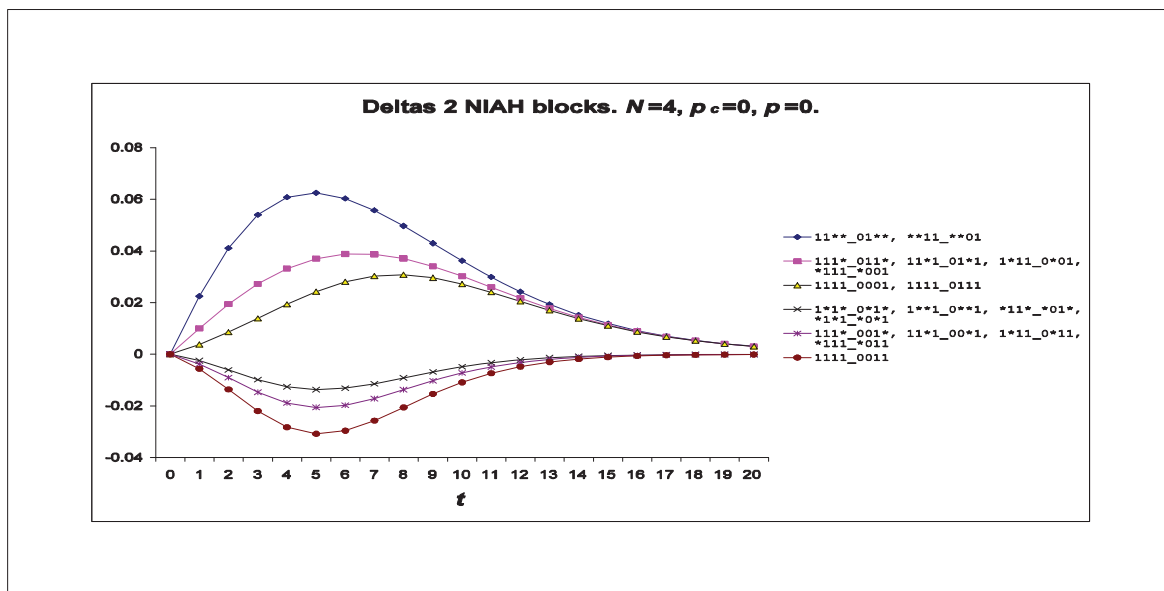


Figura 2-12 Gráfica de  $\delta_I(m,t)$  en el paisaje de 2 bloques de aguja en un pajar  $N = 4$  sin recombinación ni mutación con población inicial infinita aleatoria. La notación es  $I_m$ , es decir, el esquema y la máscara de recombinación.

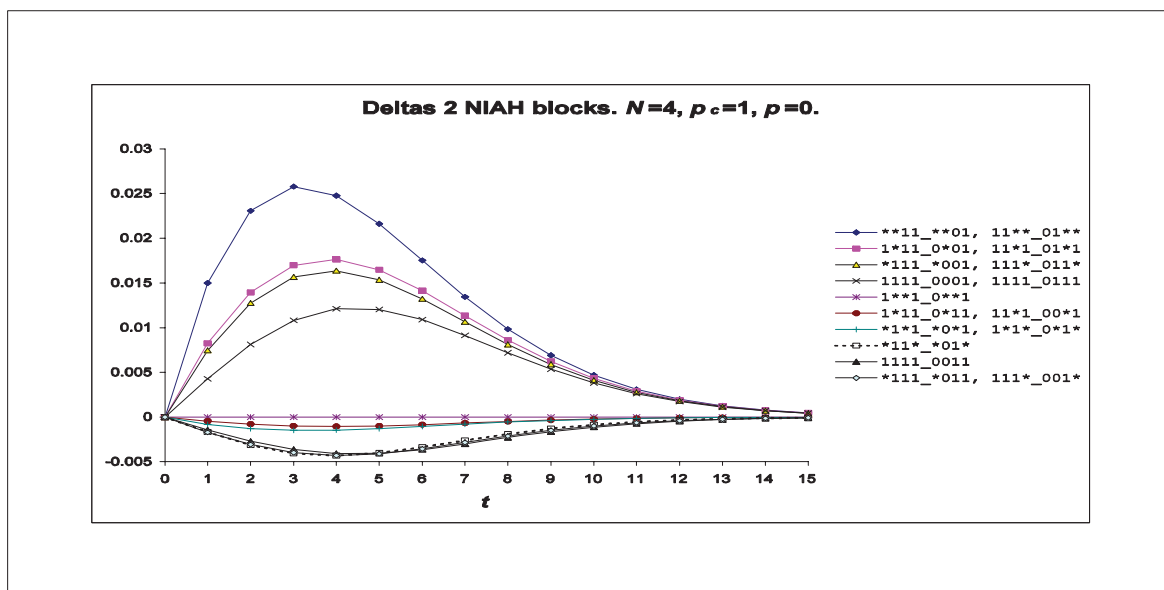


Figura 2-13 Gráfica de  $\delta_I(m,t)$  en el paisaje de 2 bloques de aguja en un pajar  $N = 4$  con recombinación sin mutación con población inicial infinita aleatoria. La notación es  $I_m$ , es decir, el esquema y la máscara de recombinación.

El GLE para determinar si la recombinación tendrá un efecto mayoritariamente reconstructivo o distributivo es aquí el grado de concordancia entre los puntos de cruzamiento y las fronteras de los bloques del paisaje. Una alta concordancia hace que se construyan más óptimos y una baja concordancia hace que los genes del óptimo (o secuencia genética dominante) sean más esparcidos entre la población.

## 2.9 Población finita con recombinación, selección y mutación

El análisis teórico de las primeras secciones de este capítulo se basa en un modelo en el que se tiene población infinita. Esto tiene la ventaja de que el modelo es determinístico pero obviamente no refleja un caso real sino el comportamiento esperado. Para completar el análisis y llevarlo a casos físicamente realizables, en esta sección se analiza la eficiencia de la búsqueda del óptimo bajo diferentes circunstancias usando poblaciones finitas. Se usan diferentes paisajes de aptitud, diferentes tasas de mutación (incluyendo la tasa óptima) y diferentes tipos de recombinación limitando los puntos de cruzamiento en los que puede ocurrir ésta. Todo esto con el fin de entender cómo interactúan los diferentes operadores genéticos entre sí y ver cuales son los GLE que emergen al actuar todos al mismo tiempo.

La métrica principal para hacer este estudio son la Curvas Características de Búsqueda descritas en el Apéndice 2.

### 2.9.1 *Aguja en un pajar*

Para tener un punto de referencia para analizar el efecto de la recombinación se muestra en la **Figura 2-14** el desempeño de un algoritmo con mutación sola sin recombinación. Nótese que el esfuerzo para encontrar la aguja (izquierda) es prácticamente el mismo con casi cualquier tasa de mutación siendo la mejor  $p = 1/2$ . En cuanto al número de generaciones necesarias para llegar al óptimo (derecha) se observa que  $p = 1/2$  es también la mejor tasa aunque la diferencia con el desempeño para tasas bajas es mucho mayor. La diferencia para tasas bajas se debe a que, para calcular el esfuerzo, solamente se cuentan las evaluaciones de la aptitud en individuos que modifican su genotipo, entonces, para tasas bajas de mutación, hay muchos individuos que no mutan y por lo tanto en cada generación se hace menos exploración requiriéndose más generaciones. Se concluye que, aunque en cuanto al esfuerzo casi cualquier tasa es igual, sí se prefiere la tasa  $1/2$  ya que con ésta se usan menos generaciones.

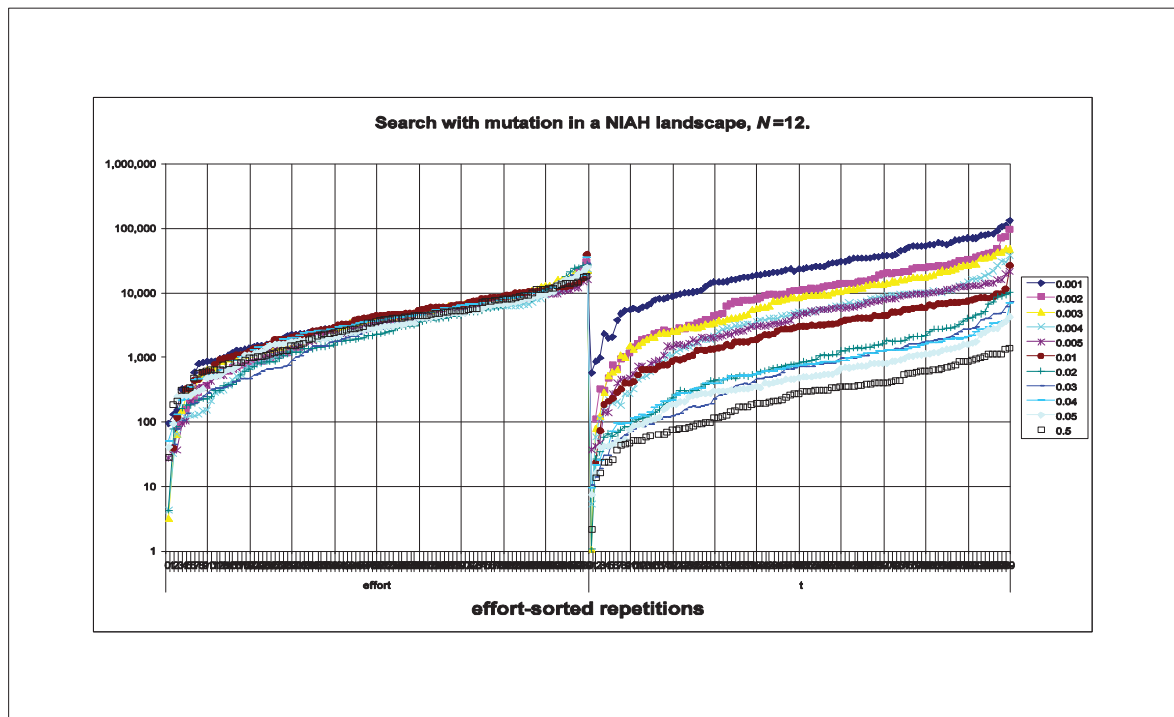
La búsqueda de la aguja únicamente mediante recombinación y sin mutación ( $p = 0$ ) será exitosa siempre y cuando se tenga inicialmente, en cada posición del genoma, por lo menos un individuo con el alelo del óptimo. Un porcentaje de las corridas no cumplirá con esta condición y el algoritmo no tendrá la posibilidad de hallar el óptimo. Es claro que mientras mayor sea el tamaño de la población esta situación será menos probable. De hecho basta con muy pocos individuos para que la probabilidad de que se de esto sea muy baja. Para el caso binario la probabilidad ? de que, en todas las posiciones del genoma, se tenga al menos un alelo igual al del óptimo es

$$1 - \frac{1}{2^s} \approx \frac{1}{2^s}$$

donde  $s$  es el tamaño de la población y  $N$  la longitud del genoma. Esta probabilidad crece con  $s$  y decrece con  $N$ , sin embargo, para valores no muy grandes de  $s$  se requieren valores sumamente grandes de  $N$  para hacer que la probabilidad  $\frac{1}{2^s}$  decrezca un poco. Por ejemplo, para  $N = 10000$  y  $s = 30$  se tiene que  $\frac{1}{2^s} = 0.999991$ . Es decir, que la probabilidad de que no se tenga esta situación es tan baja que no debe considerarse demasiado.

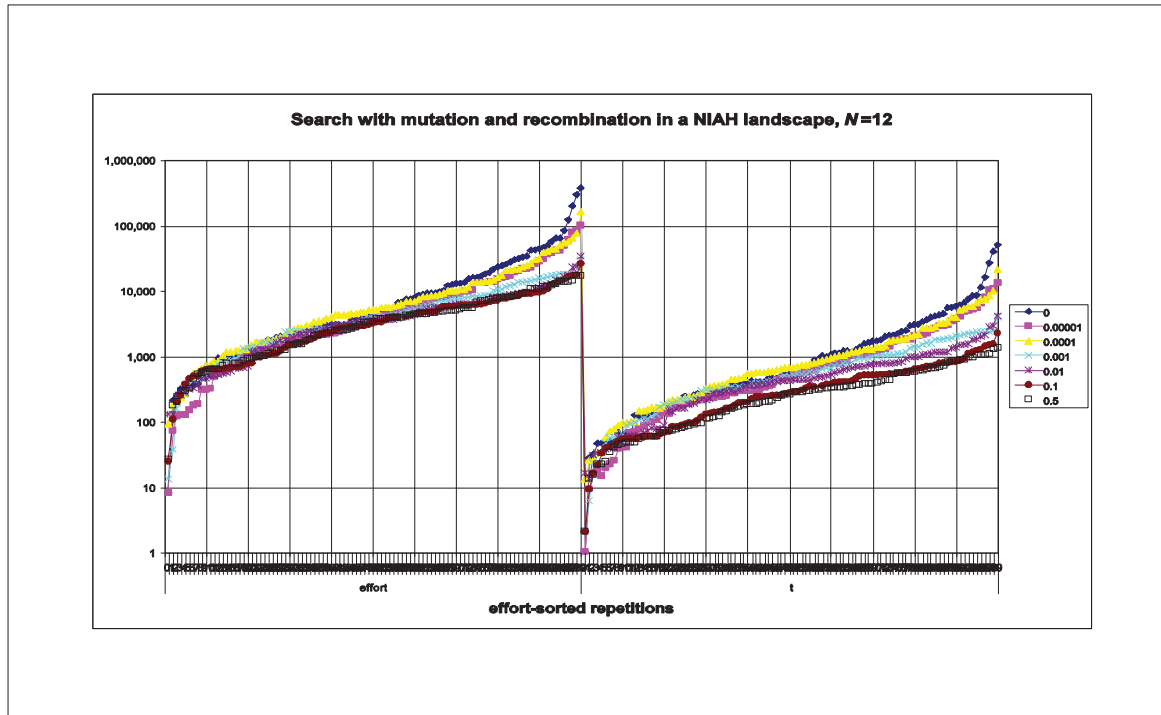
Para poder hacer una búsqueda equilibrada en el espacio de los genomas, es necesario que se tenga, para cada posición del genoma, la mitad de individuos con cero y la mitad con uno (suponiendo claro que no se sabe la configuración de la aguja). La situación más difícil para encontrar el óptimo, sin que sea imposible, es cuando en la población hay pocas coincidencias con el óptimo en muchas posiciones del genoma.

Al incorporar mutación al algoritmo es posible salvar la falta de ceros o unos (coincidencias con el óptimo) en alguna posición. Dado que el paisaje de aptitud es plano mientras no se encuentre el óptimo, la mutación llevará a la población gradualmente a una distribución uniforme de los individuos en el espacio de los genotipos (el centro del simplex) en donde se cumple la condición ideal para la búsqueda por recombinación, entonces, aunque la tasa de mutación sea baja, eventualmente se llegará a esa situación y la recombinación hará el resto del trabajo.



**Figura 2-14** Curvas Características de Búsqueda de esfuerzo y de tiempo con mutación sola en el paisaje de aguja en un pajar para varias tasas de mutación.  $N = 12$ .

En la **Figura 2-15** se muestra el desempeño de un algoritmo con recombinación para varias tasas de mutación. Puede verse (izquierda) que para una tasa de mutación cero la búsqueda es ligeramente más lenta debido a que la población no se acerca a la distribución uniforme. Nótese también que basta con tasas de mutación muy pequeñas para que la búsqueda sea muy similar a la búsqueda aleatoria.



**Figura 2-15** Curvas Características de Búsqueda de esfuerzo y de tiempo con recombinación y mutación en el paisaje de aguja en un pajar para varias tasas de mutación.  $N = 12$ .

Cuando la tasa de mutación es muy baja el grueso del trabajo lo hace la recombinación y el resultado es el mismo en cuanto a esfuerzo (izquierda) que con búsqueda aleatoria pero no en cuanto a tiempo (derecha). Esto es notable al comparar la **Figura 2-15** con la **Figura 2-14**. Véase que el tiempo para llegar al óptimo (derecha) es mucho mayor si se usa solamente mutación baja (**Figura 2-14**) que con mutación y recombinación (**Figura 2-15**).

En resumen, para el paisaje de aguja en un pajar, la búsqueda no puede ser mejor que la búsqueda aleatoria, la mutación es necesaria y con tasas muy bajas es suficiente para obtener un buen desempeño aunque tasas mayores son ligeramente mejores.

Por lo anterior se deduce entonces que la tasa de mutación óptima en este paisaje de aptitud es  $\frac{1}{2}$  ya que, con esta tasa, la población llega a una distribución uniforme en una sola generación. Lo único que requiere el operador de recombinación para ser un buen operador de búsqueda es diversidad genética.

## 2.9.2 Concatenación de paisajes de Aguja en un pajar

Ya se vio que en paisaje de aguja en un pajar la recombinación no ayuda más allá que como un operador de búsqueda aleatoria (cuando la población es aleatoria), entonces ¿Cuál debe ser la característica del paisaje para que la recombinación ayude más allá que como operador de búsqueda aleatoria? En la parte teórica de este capítulo se encontró que en paisajes modulares la recombinación parece tener efectos positivos en la concentración del óptimo. La modularidad se da cuando secciones del genoma evolucionan independientemente una de la otra. En estos casos la recombinación debería unir soluciones parciales en una sola que contenga a ambas produciéndose así la evolución hacia algo mejor. Sin embargo se han publicado resultados contrarios a esta aseveración como por ejemplo en el caso de los caminos reales o “Royal Roads” **¡Error! No se encuentra el origen de la referencia.** ]. Para aclarar esto se usa aquí un paisaje con bloques concatenados cada uno con un paisaje de aptitud de aguja en un pajar.

### 2.9.2.1 Concatenación de 3 paisajes de Aguja en un pajar de 4 bits cada uno

En la **Figura 2-16** se muestran los resultados de la búsqueda en un paisaje con 3 bloques concatenados de aguja en un pajar, cada uno con 4 bits. En total 12 bits. Se usó recombinación en 2 puntos con una probabilidad de cruzamiento de 0.9 y con una población de 13 individuos inicializados aleatoriamente. Se muestran 5 gráficas, con 2, 3, 4, 6 y 12 puntos de cruzamiento para la recombinación distribuidos uniformemente a lo largo de los 12 bits. Esto es, por ejemplo, para 3 puntos de cruzamiento se permite que la recombinación sea únicamente en los puntos 0, 4 y 8 que, en este caso, coinciden precisamente con las fronteras de los 3 bloques que forman el paisaje de aptitud.

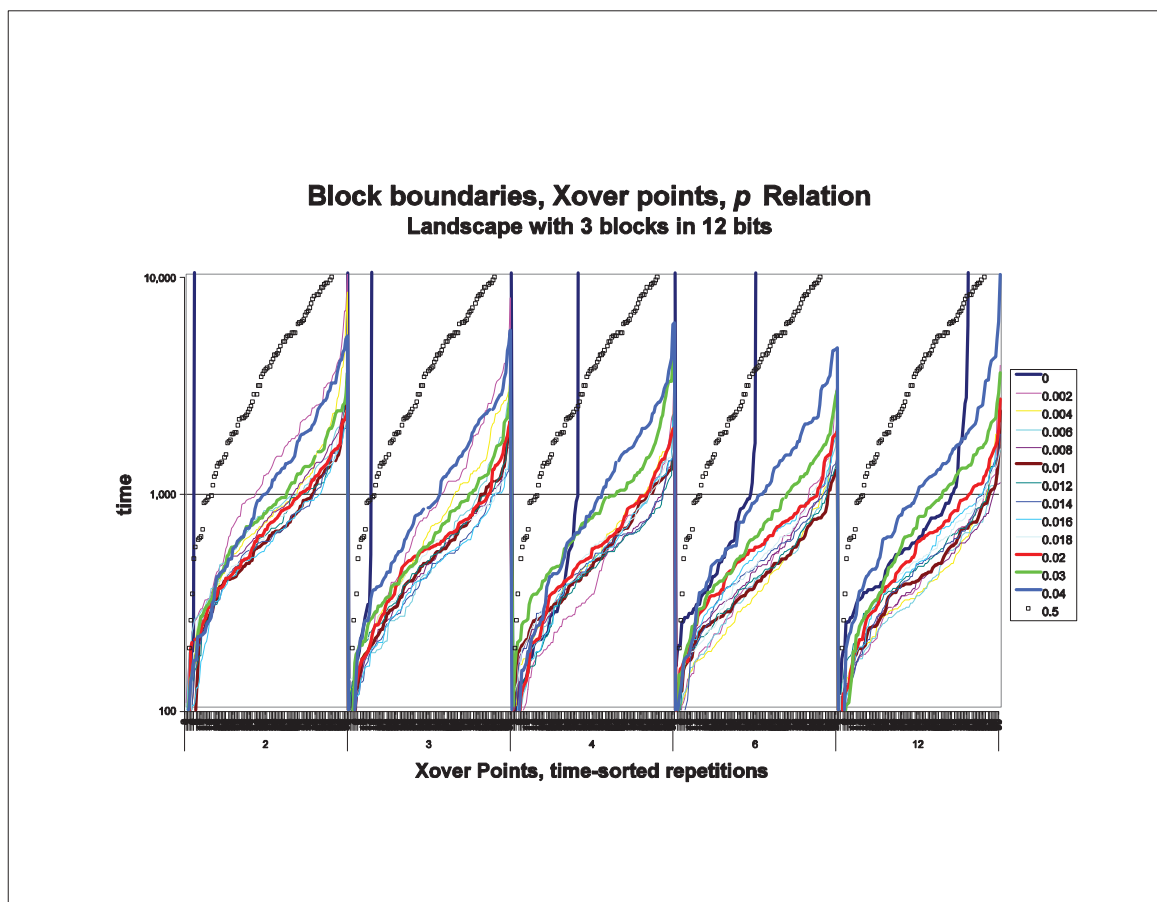
Para el caso de 12 puntos de cruzamiento véase en la **Figura 2-16** (extrema derecha) que sin mutación (línea gruesa azul oscuro) la búsqueda es bastante buena en aproximadamente el 80% de los casos ya que la curva está por debajo de la búsqueda aleatoria (cuadrados negros) pero en el otro 20% es muy mala ya que no fue posible encontrar el óptimo en el lapso de tiempo (evaluaciones) máximo de cada corrida. Lo anterior se debe a que en un porcentaje de las corridas (20%) se da el caso de que en alguna posición del genoma todos los individuos difieren con el óptimo lo cual hace imposible que la recombinación lleve al óptimo. Al aumentar la tasa de mutación la situación mejora considerablemente. Al seguir aumentando la tasa de mutación se llega a una tasa ideal (muy baja). Para tasas de mutación mayores a esta tasa ideal la búsqueda pierde eficiencia pareciéndose cada vez más a la búsqueda aleatoria debido a la destrucción por mutación de los óptimos de bloque que ya fueron encontrados impidiendo su explotación.

En el caso de 6 puntos de cruzamiento se tienen menos puntos de cruzamiento “destructivos”, es decir, puntos fuera de las fronteras entre los bloques del paisaje de aptitud y se conservan los puntos de cruzamiento en dichas fronteras. Esto hace que los óptimos de cada bloque sean menos destruidos por la recombinación. Sin embargo, esto también limita la capacidad de exploración de las distintas combinaciones genéticas lo cual hace necesario usar tasas de mutación ideales ligeramente mayores que en el caso anterior. Al usar tasas de mutación mayores se tiene una mayor probabilidad de destrucción (por mutación) de los óptimos que ya

fueron encontrados en los bloques haciendo (ligeramente) más lenta la búsqueda.

Cuando se usan 4 puntos de cruzamiento toda recombinación resulta destructiva (sin dejar de ser explorativa) ya que estos puntos no coinciden con las fronteras de los bloques del paisaje. De nuevo aquí, al tener menos puntos de cruzamiento se limita aún más la capacidad de exploración de la recombinación haciéndose necesarias tasas de mutación mayores.

Con 3 puntos de cruzamiento éstos coinciden con las fronteras de los bloques. Podría pensarse que este sería el mejor caso por tratarse de puntos de cruce que no destruyen los óptimos de bloque ya encontrados pero sorprendentemente no es así. La razón es que la búsqueda dentro de los bloques debe hacerse con tasas de mutación mayores que en los casos anteriores lo cual tiende a destruir los óptimos ya encontrados y resulta muy difícil su recuperación.



**Figura 2-16** Curvas Características de la búsqueda mediante mutación, recombinación anular de 2 puntos restringida a 2, 3, 4, 6 y 12 posiciones y selección en un paisaje de aptitud con 3 bloques de aguja en un pajar de 4 bits,  $p_c = 0.9$ , población de 13 individuos y diversas tasas de mutación.

Lo mismo sucede cuando se usan 2 puntos de cruce y es por eso que este caso es el menos eficiente para hallar el óptimo global.

En todos los casos la búsqueda es mejor que la búsqueda aleatoria debido a que, si la mutación no interfiere, siempre se aprovecha cuando algún individuo consigue el óptimo en alguno de los bloques o módulos. Al pasar algún individuo por un óptimo de un bloque, su configuración genética será favorecida por tener mayor aptitud reflejándose en un mayor número de alelos de ese óptimo en la población. Esta ventaja se mantiene aún cuando ese óptimo sea destruido por medio de la recombinación ya que el paisaje (en el bloque) en ese caso volvería a ser plano, y solamente la mutación puede eliminar dicha ventaja. Esta ventaja tiene como consecuencia que dicho óptimo de bloque sea visitado con mayor frecuencia generándose con cada visita una mayor ventaja. Esto explica porqué el mejor caso es en donde se tienen 12 puntos de cruzamiento ya que, así, la búsqueda de nuevas combinaciones genéticas se hace principalmente con la recombinación y se hace uso mínimo de la mutación la cual hace que en más ocasiones se pierda el óptimo en los bloques de manera irreversible (o muy difícilmente reversible), que no es el caso con la recombinación (cuando el paisaje es plano).

### 2.9.2.2 *Paisaje con N bloques concatenados de aguja en un pajar*

En este paisaje cada bloque tiene longitud 1 así que cuando en alguna posición del genoma el alelo coincide con el alelo del óptimo ya se tiene un bloque completo. En este caso la recombinación no puede destruir ningún bloque ya encontrado. Lo anterior implica que no es necesaria la mutación ya que ésta sí puede destruir bloques, por lo tanto es mejor hacer la búsqueda por recombinación. Solamente es necesario que en la población inicial se tenga para cada posición del genoma por lo menos un individuo con el alelo correspondiente del óptimo.

En la **Figura 2-17** se muestra cómo el mejor caso es con tasa de mutación 0, lo que significa que la mutación solamente hace menos eficiente la búsqueda y que la recombinación es el principal operador en la búsqueda (en este caso la búsqueda aleatoria es mucho menos eficiente y por esto no se muestra en la gráfica).

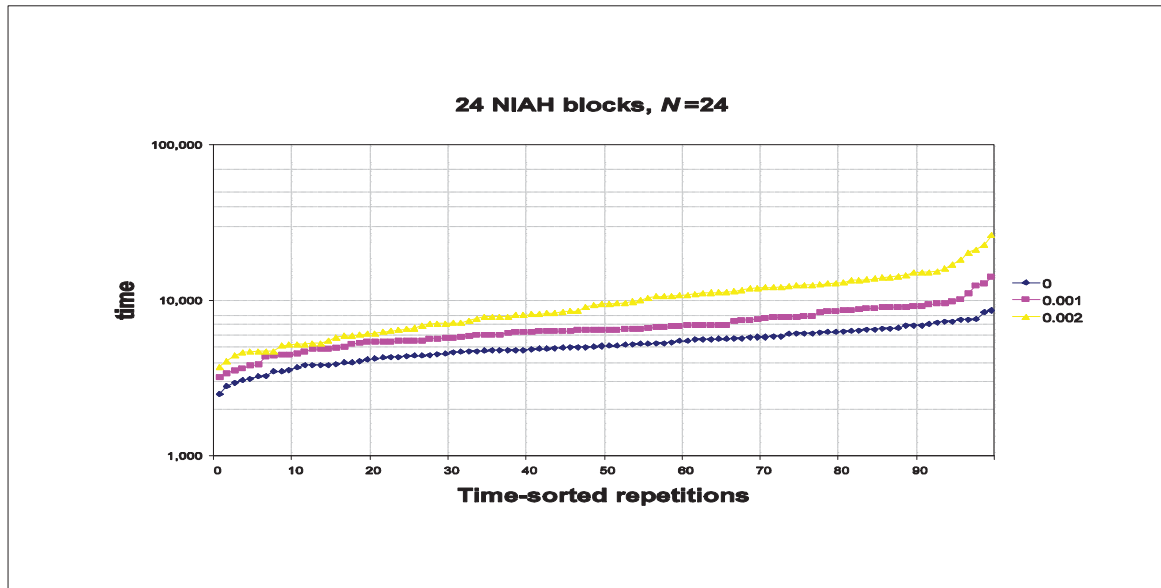
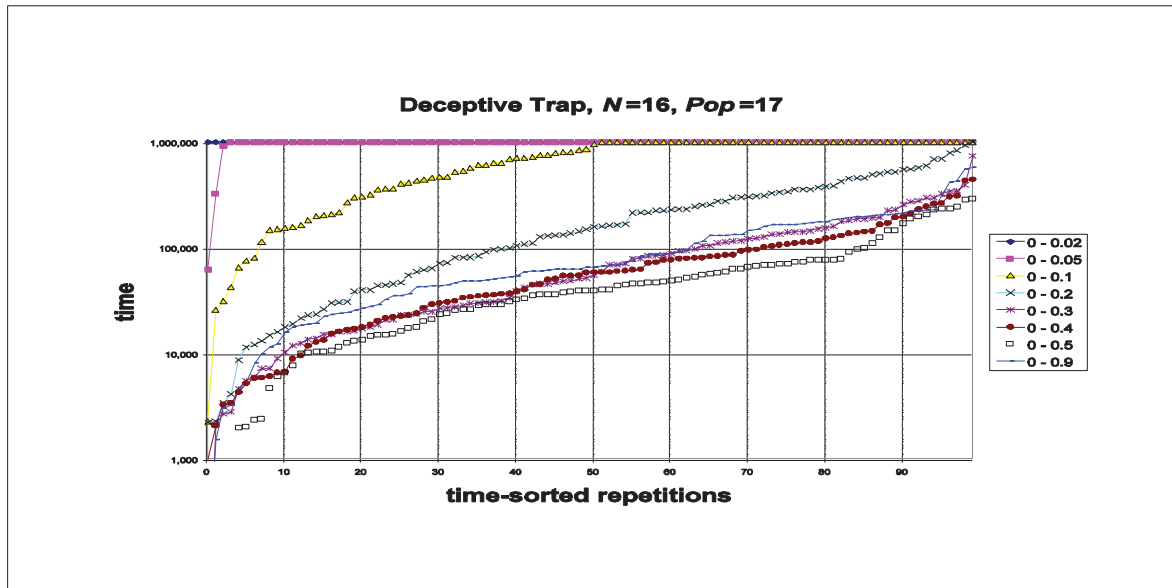


Figura 2-17 Curvas características de búsqueda con recombinación para tasas de mutación 0, 0.001 y 0.002 en el paisaje de  $N$  bloques de aguja en un pajar.  $N = 24$ .

### 2.9.3 Trampa o paisaje engañoso

El paisaje anterior resultó “ideal” para ser resuelto mediante recombinación ya que el operador lleva a la población hacia el óptimo rápidamente. Ahora se tiene un paisaje que es exactamente el mismo que el anterior solo que el óptimo se encuentra en el extremo opuesto, si el óptimo era antes la secuencia de puros unos ahora es la que tiene solo ceros. En este caso entonces, mientras *menos* coincidencias se tengan con el óptimo mayor será la aptitud de los individuos lo cual lo hace un paisaje engañoso y la recombinación debe ayudar a alejar a la población del óptimo.





**Figura 2-18** Curvas características de búsqueda para diversas tasas de mutación en el paisaje de trampa o engañoso.

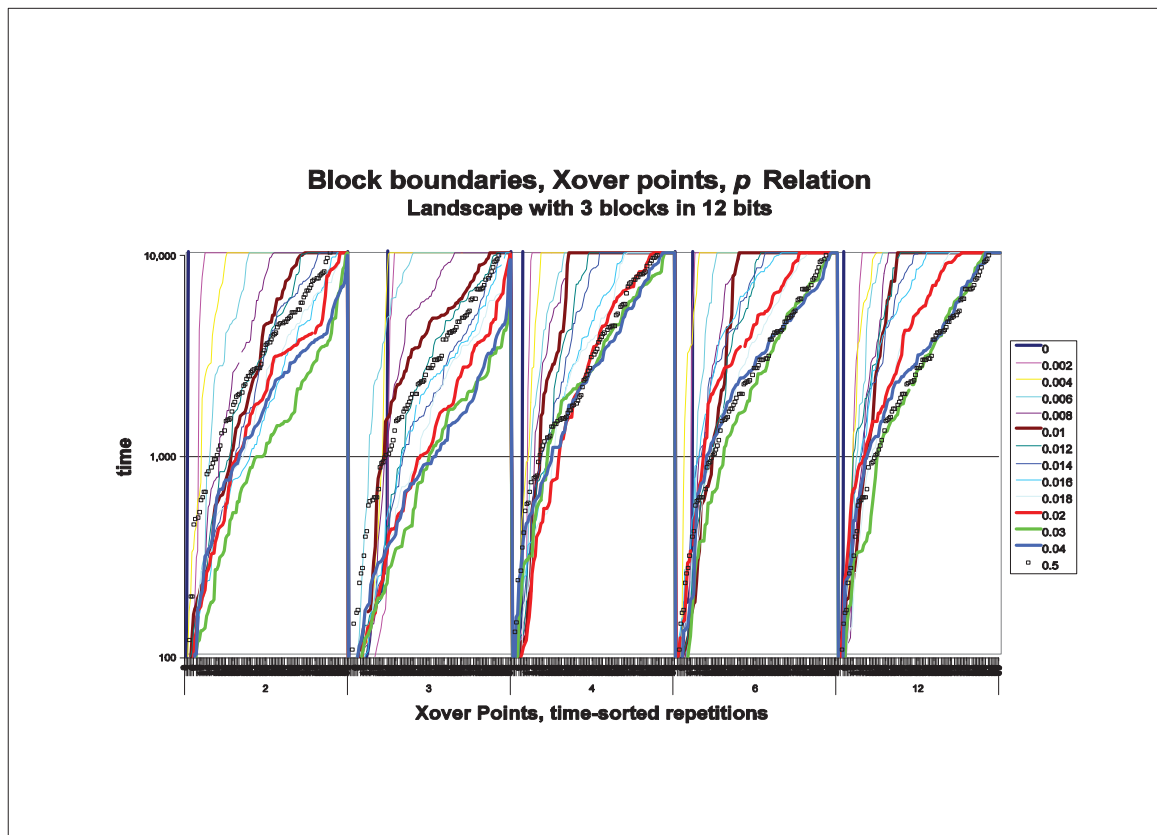
La población tiende formar una cuasiespecie (a agruparse) alrededor del genotipo cuyos genes son totalmente discordantes con los del óptimo, al cual es un subóptimo local. Se hace necesaria una tasa de mutación mayor para extender el alcance de la cuasiespecie y que sea más probable modificar los genes de los individuos para que coincidan con los del óptimo. Esto es evidente en la **Figura 2-18** en donde se ve que el mejor caso es cuando la tasa de mutación es 0.5, es decir, búsqueda aleatoria.

Puede pensarse que con tasas de mutación cercanas a uno podría hallarse el óptimo más rápidamente ya que el óptimo está situado en el extremo opuesto al lugar al que tiende la población siendo necesario mutar casi todos o todos los bits del genotipo dominante (el subóptimo). Esto no es así porque, al usar por ejemplo  $p = 1$  (o alguna tasa muy cercana a 1), la población oscila entre su estado actual y el estado opuesto. En estos estados opuestos, el paisaje de aptitud favorece a aquellos genotipos que están más lejos del óptimo, que son, alternadamente, unos en un estado y otros en el otro estado ya que invierten su distancia con el óptimo. Como la intensidad de dicha preferencia es la misma en ambos estados, ésta queda cancelada efectivamente y no hay un avance real hacia ningún lado, es decir, que la población no puede acercarse ni al óptimo ni al subóptimo. Véase el caso cuando la señal engañosa no es igualmente fuerte en todo el paisaje (2.9.5 más adelante).

### 2.9.4 *Trampas concatenadas*

Como se ha visto, la recombinación es buena en paisajes con modularidad (hacia el óptimo) y mala en paisajes engañosos en los que hay una modularidad inversa, es decir, en dirección contraria al óptimo. Entonces ¿Qué pasa si el paisaje tiene ambas propiedades? Un paisaje así es uno que es modular pero que en cada módulo es engañoso.

Aquí se define un número de bloques donde cada bloque es un paisaje de trampa. En este paisaje hay modularidad local engañosa pero también hay una modularidad global de bloques hacia el óptimo. Entonces se da un compromiso entre las dos. Puede ser que lo engañoso del paisaje de cada bloque sea más fuerte que la modularidad de bloques o viceversa. La recombinación siempre ayuda a subir por paisajes modulares y, como cada bloque es modular pero en sentido hacia el subóptimo, la recombinación resulta perjudicial, sin embargo, también es benéfica a nivel de bloques. Por esto, la forma más eficiente de llegar al óptimo global es que la mutación se encargue de hallar los óptimos de bloque y la recombinación de unir éstos.



**Figura 2-19** Curvas características de búsqueda con diferentes tasas de mutación y diferente número de puntos de recombinación en el paisaje de trampas concatenadas (3 bloques de 4 bits,  $N = 12$ ).

En la **Figura 2-19** se muestra que, para cruzamiento en cualquiera de 12 puntos (con  $N = 12$ ) el mejor caso es búsqueda aleatoria ( $p = 0.5$ ) ya que la recombinación destruye los óptimos de bloque ya encontrados y lleva a la población hacia el subóptimo de cada bloque haciendo necesarias tasas de mutación altas que también destruyen los óptimos de bloque y resulta que la única manera de hallar el óptimo global es hallando todos los óptimos de bloque al mismo tiempo. Cuando se limitan los puntos de cruzamiento a los puntos en donde están las fronteras entre los bloques (segundo caso de izquierda a derecha), es decir, a 3 puntos, se tiene la mejor eficiencia posible ya que la mutación se encarga de encontrar los óptimos de bloque con tasas de mutación relativamente bajas permitiendo que, al hallarse uno, éste domine en la población y que no se extinga ni por mutaciones ni por recombinaciones. Cuando algún otro individuo

encuentra un óptimo de bloque, la recombinación hace más probable que estos óptimos se unan en un individuo acelerando el proceso evolutivo.

En un paisaje donde los bloques son cortos la tasa de mutación ideal es baja ya que eso es suficiente para explorar en estos bloques. Si los bloques son largos lo contrario aplica.

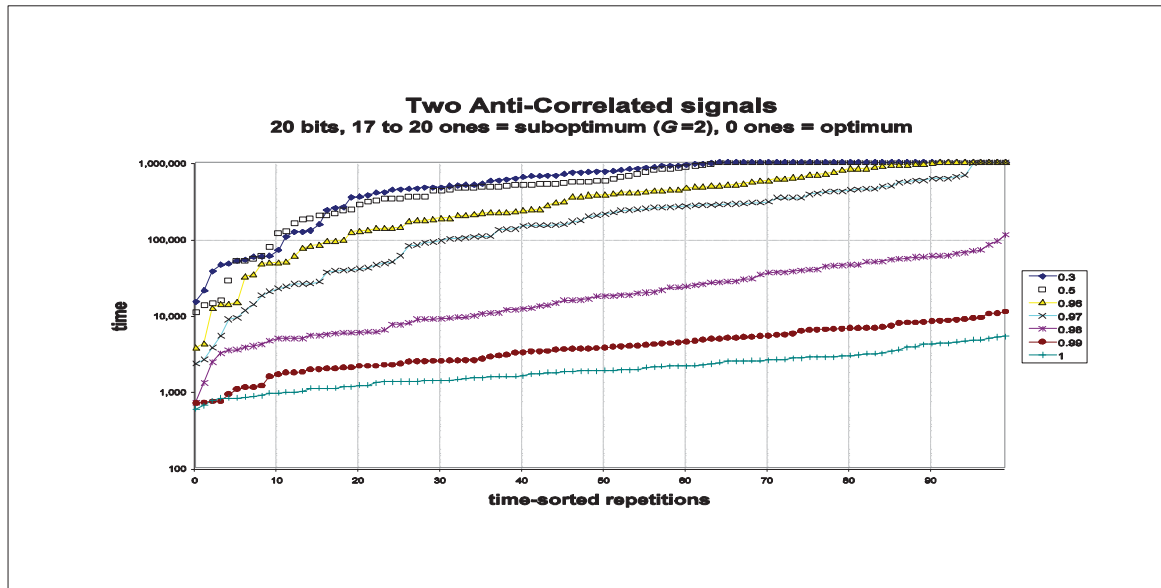
### 2.9.5 *Dos señales anticorrelacionadas*

En problemas reales, no todas las señales engañosas tienen la misma forma y de hecho son de muy diversas formas. Para ver que pasa con un paisaje engañoso diferente al anterior se usa ahora un paisaje que tiene dos señales en forma de agujas. Una de ellas es el óptimo global y es una sola secuencia genética, la otra señal está situada en el extremo opuesto al óptimo y es una señal ancha o degenerada, es decir, que consta de más de una secuencia genética. El resto del espacio de búsqueda es plano. En el ejemplo que se muestra en la **Figura 2-20** el óptimo global es la secuencia con 20 ceros y el subóptimo consta de las secuencias que tienen de 17 a 20 unos.

En cierta forma este paisaje se parece al paisaje de trampa en el sentido de que hay una señal que aleja a la población del óptimo.

Como puede apreciarse en la gráfica, la mejor tasa de mutación posible es 1. Esto se debe a lo siguiente: Es más probable que la población encuentre la señal engañosa primero que la señal del óptimo por ser la primera más ancha impulsándose la proliferación de las secuencias con muchos unos y, por ende, la secuencia con todos los unos. Es decir que la población se siente atraída por la señal del subóptimo. Con una tasa de mutación 1, la población está siempre alternando entre la actual y la población opuesta, entonces, cuando la población es atraída al subóptimo, en la siguiente generación se va a encontrar más cerca del óptimo.

En general, para paisajes con señales engañosas opuestas al óptimo, la tasa de mutación  $p = 1$  provocará que, si la intensidad con que la población es atraída hacia el opuesto del óptimo (cerca de éste) es mayor que la intensidad con que es repelida del óptimo (cerca de éste), la población se irá acercando poco a poco hacia el opuesto del óptimo (alternadamente cada 2 generaciones). Cuando la secuencia opuesta al óptimo es encontrada, en la siguiente generación se habrá encontrado el óptimo. Entonces se tiene que la señal engañosa realmente ayuda en la búsqueda pero solamente cuando  $p = 1$  o tiene un valor cercano a 1.



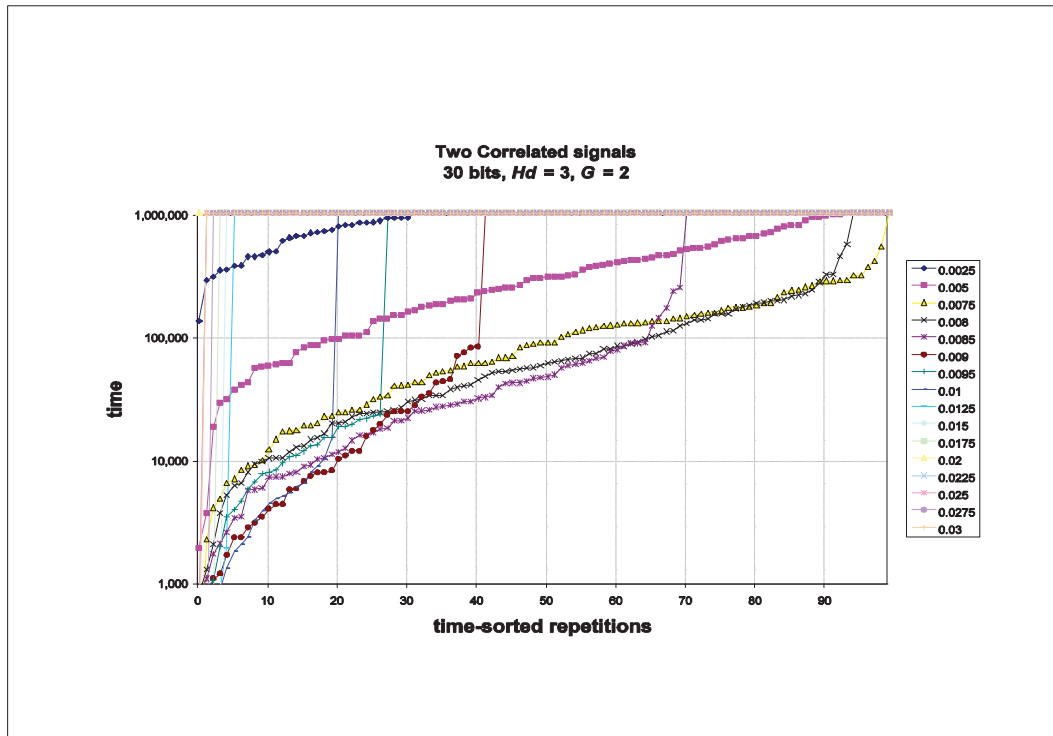
**Figura 2-20** Curvas características de búsqueda con diferentes tasas de mutación en el paisaje de dos señales anticorrelacionadas. 20 ceros es el óptimo y de 17 a 20 unos es el subóptimo que tiene una  $G = 2$ .

### 2.9.6 Dos señales correlacionadas

En el **¡Error! No se encuentra el origen de la referencia.** en la sección **¡Error! No se encuentra el origen de la referencia.** se hizo búsqueda (sin recombinación) en paisajes con dos señales, en los que una de ellas es el óptimo, a diferentes distancias de Hamming una de la otra. Se analizó en qué manera influye la presencia de la señal del subóptimo cuando está cerca o lejos del óptimo. Se usan ahora estas mismas pruebas pero agregando recombinación para ver si hay cambios cualitativos en los resultados.

En este paisaje se tienen dos señales: el óptimo y otra señal muy cercana al óptimo pero separada de éste siendo el resto del paisaje plano. La idea es analizar el comportamiento del sistema en el caso cuando la población queda varada en el óptimo local y debe desde allí encontrar el óptimo.

Si la tasa de mutación es muy baja habrá muy poca exploración del espacio y la búsqueda será lenta. Si al contrario la tasa de mutación es demasiado alta se perderá la señal del subóptimo y consecuentemente la búsqueda pierde orden y resulta equivalente a una búsqueda aleatoria. Esto es evidente en la **Figura 2-21** donde se muestran curvas características de búsqueda cuando todos los individuos se inicializan en el subóptimo cuya  $G = 2$  y el óptimo está a una distancia de Hamming 3 del subóptimo. Las curvas correspondientes a las tasas de mutación más altas se doblan hacia arriba abruptamente indicando que hay un número de repeticiones que pierden el subóptimo y consecuentemente la búsqueda se convierte en búsqueda aleatoria. La tasa de mutación ideal es la mayor entre las que conservan el subóptimo.

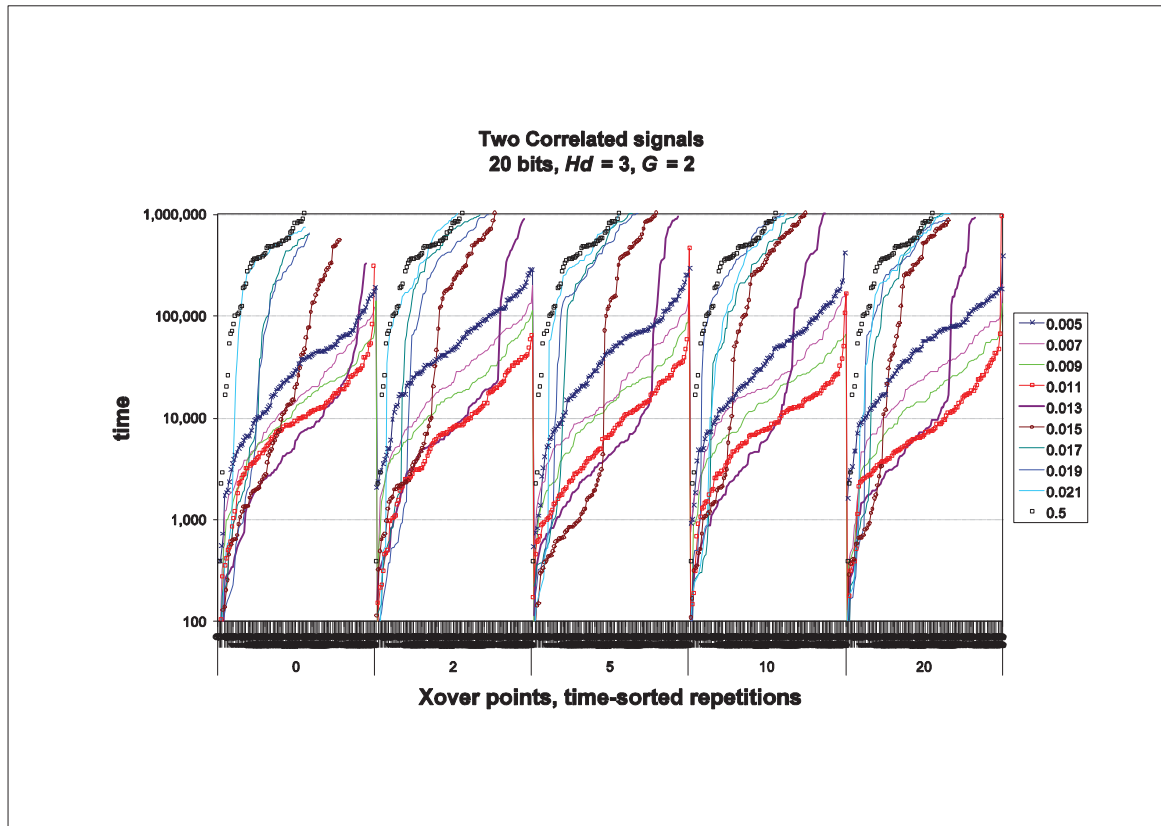


**Figura 2-21** Curvas características de búsqueda con diferentes tasas de mutación a partir de una población inicial localizada en un óptimo local con  $G = 2$  con el óptimo a una distancia de Hamming 3.  $N = 20$ .

Este mismo efecto puede verse en el capítulo 1 en donde no se aplica recombinación. Para tasas de mutación muy bajas la aplicación de recombinación no produce efecto importante pero para tasas de mutación cercanas a la tasa crítica de la señal dominante sí ya que hay más exploración y se reduce efectivamente el valor de la tasa crítica de mutación.

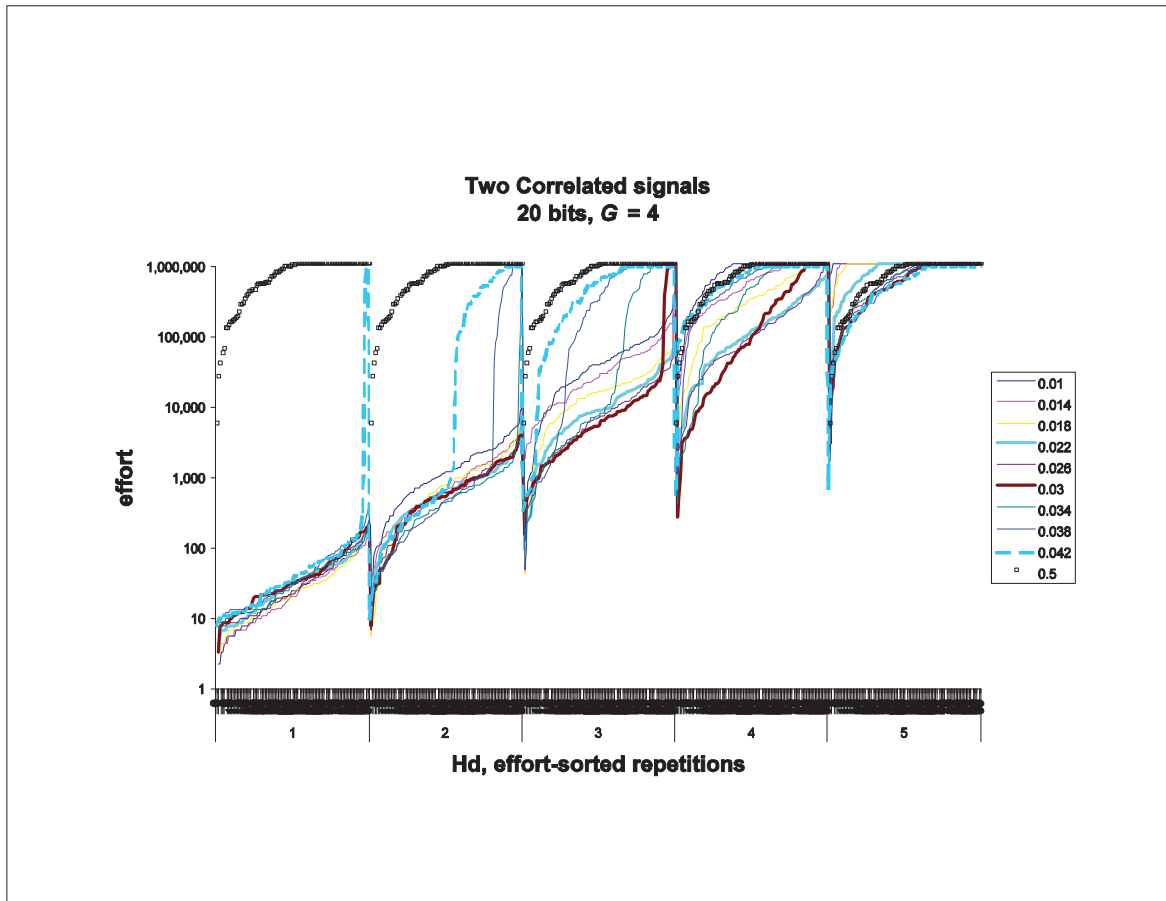
Como puede verse en la **Figura 2-22**, la tasa de mutación ideal para salir del óptimo local y llegar al óptimo depende muy poco del número de puntos permitidos para la recombinación (a menos que existan rasgos de modularidad u otras señales en el paisaje que desvíen la búsqueda en cuyo caso la recombinación sí afectaría, los resultados mostrados aquí son con el resto del paisaje plano). Esta tasa depende principalmente de la fuerza de la señal del subóptimo aunque sí se nota una ligera reducción al usar más puntos de cruzamiento. Principalmente, la tasa ideal está relacionada con la tasa crítica de la señal del subóptimo.

Esta tasa ideal para hallar el óptimo tampoco depende mucho de la distancia a la que se encuentre el óptimo del subóptimo, pero, cuando el óptimo está fuera del alcance de la cuasiespecie alrededor del subóptimo, entonces la búsqueda debe hacerse con tasas de mutación superiores a la tasa crítica del subóptimo lo que implica que la búsqueda ya no estará localizada cerca del subóptimo sino esparcida en todo el espacio de búsqueda y como el resto del paisaje es plano se tendrá búsqueda aleatoria.



**Figura 2-22** Curvas características de búsqueda con diferentes tasas de mutación a partir de una población inicial localizada en un óptimo local con  $G = 2$ , con el óptimo a una distancia de Hamming 3, usando recombinación con un número de puntos de cruzamiento limitado a 0, 2, 5, 10 y 20.  $N = 20$ .

En la **Figura 2-23** se muestran las curvas características de búsqueda con distintas distancias de Hamming entre el óptimo y el subóptimo. Obsérvese que para una distancia de Hamming de 5 (a la derecha) la búsqueda ya no puede ser mejor que búsqueda aleatoria pero sí peor para tasas de mutación muy bajas. Puede verse que la tasa 0.03 (línea café gruesa) es buena para cualquier distancia de Hamming mientras que las tasas menores solo son buenas para distancias de Hamming bajas. Aquí, entonces, se puede decir que la tasa ideal está relacionada con la tasa crítica del subóptimo y que esta es mucho menor que la tasa crítica teórica  $\ln(4)/20 = 0.069$



**Figura 2-23** Curvas características de búsqueda con diferentes tasas de mutación a partir de una población inicial localizada en un óptimo local con  $G = 4$  con el óptimo a una distancia de Hamming de 1, 2, 3, 4 y 5 con una probabilidad de cruzamiento de 0.9.  $N = 20$ .

## 2.10 Conclusiones

En este capítulo se hizo un análisis de los efectos que tiene la recombinación en un sistema genético bajo diferentes circunstancias mediante herramientas tanto teóricas como empíricas. Se pudo lograr clarificar el papel que juega la recombinación por sí sola y la forma en que interactúa con los operadores de selección y mutación al descubrir los grados de libertad efectivos relacionados con su aplicación. Se pudo entonces determinar bajo qué circunstancias la recombinación es benéfica para la búsqueda genética y bajo cuales no.

La primera conclusión es que el efecto de la recombinación es variable. Se puede tener desde un efecto casi nulo hasta un efecto equivalente a la búsqueda aleatoria. El grado de libertad efectivo para que se dé uno u otro caso es el grado de dominio de un solo genotipo sobre el resto de la población y el grado de diversidad genética en la misma. Con alto dominio de un genotipo o baja diversidad genética el efecto de la recombinación es casi nulo. Si hay una alta



diversidad genética y no hay una ventaja de ningún genotipo la población está cerca del centro del simplex y entonces la recombinación tiene un efecto equivalente a una búsqueda aleatoria. El efecto es más importante, entonces, cuando hay diversidad genética pero hay también algún genotipo (o clase de genotipos) con un grado de dominio no despreciable pero tampoco muy alto. En este caso, la recombinación hace que la genética dominante se mezcle con el resto produciéndose combinaciones mayoritariamente cercanas al genotipo dominante pero con variantes distintas. Las frecuencias de cada genotipo están dadas por la variedad de Geiringer **¡Error! No se encuentra el origen de la referencia. ]**.

Cuando se aplica mutación, si ningún genotipo de los que están presentes en la población tiene una ventaja reproductiva, se llega pronto (en pocas generaciones) al centro del simplex en donde la diversidad genética es máxima y ningún genotipo domina, por lo tanto, la recombinación tiene un efecto de búsqueda aleatoria.

Si en alguna generación se llega a hallar algún genotipo con una aptitud ventajosa sobre los demás y no hay mutación, la población tenderá a converger totalmente hacia ese genotipo. Con mutación, el grado de convergencia lo determina la tasa de mutación y la ventaja del genotipo hallado. La recombinación no influye en este grado de convergencia ya que el resto de los genotipos tienen un mismo nivel de aptitud. En otras palabras, cuando el resto del paisaje es plano, el alcance de la cuasiespecie emergente alrededor del genotipo con ventaja de aptitud no depende de la recombinación.

La modularidad del paisaje es un grado de libertad efectivo cuando se tiene recombinación. Los paisajes de aptitud pueden presentar modularidad hacia el óptimo o en otras direcciones en diferentes regiones y a diferentes niveles. La recombinación hace más intensa la tendencia de la población a seguir la dirección de las modularidades del paisaje, por lo que su efecto depende directamente de éstas. Mientras más modular hacia el óptimo sea el paisaje más beneficio da la recombinación y viceversa.

Para que la población pueda salir de óptimos locales se requiere usar una tasa de mutación suficientemente alta. Si no hay modularidad hacia el óptimo local la recombinación hace que se requieran tasas de mutación menores. Si la modularidad es hacia el óptimo local la recombinación hace que sean necesarias tasas de mutación mayores. Si la tasa de mutación es muy alta entonces es posible salir de cualquier óptimo local con el consecuente riesgo de perder toda información obtenida previamente haciéndose la búsqueda equivalente a búsqueda aleatoria.

Otro grado de libertad efectivo es el grado de concordancia entre el conjunto de posibles puntos de cruzamiento para la recombinación y los bordes de los módulos o bloques en el paisaje de aptitud. A mayor concordancia entre estos mayor será el efecto de refuerzo a la modularidad del paisaje por parte de la recombinación.



## Notas

## Capítulo 3 **Operadores genéticos dependientes del Ranqueo**

En los capítulos anteriores se ha logrado adquirir un entendimiento de la dinámica en un algoritmo genético simple (SGA) mediante la identificación de los grados de libertad efectivos presentes en él. Esto resulta útil en el área de genética poblacional para entender mejor la evolución natural de las especies. Por el lado computacional, este conocimiento tiene implicaciones en aplicaciones prácticas de los algoritmos genéticos. Cuando un AG es usado como herramienta de búsqueda, al entender la dinámica presente, es posible ajustarlo y/o modificarlo para mejorar su eficiencia y eficacia.

Se ha visto también, entre otras cosas, cómo es que el tener información acerca del paisaje de aptitud y del estado de la población puede ayudar a seleccionar una tasa de mutación ideal o las máscaras de recombinación más útiles. Tasas bajas son recomendables cuando en el paisaje hay subóptimos locales débiles o cercanos al óptimo global y tasas altas son recomendables si hay subóptimos locales robustos o lejanos al óptimo global. Los puntos de cruzamiento para la recombinación que son más útiles son aquellos que coinciden con las fronteras entre los bloques o módulos del paisaje de aptitud.

Sin embargo, es común que no se tenga conocimiento de todos los detalles del paisaje de aptitud, de hecho, lo normal es que se sepa muy poco acerca de ellos. A este tipo de problemas se les conoce como de “caja negra”. Este tipo de problemas son la razón por la que se usan los algoritmos genéticos (y otras heurísticas). Si se supieran todos los detalles del paisaje de aptitud ya no tendría sentido hacer siquiera una búsqueda. Hay incluso muchos problemas reales en los que el paisaje de aptitud cambia con el tiempo. Bajo este escenario incierto, se requieren entonces algoritmos robustos ante cualquier eventualidad. El algoritmo genético simple (SGA) no puede ser eficiente para todos los casos y requiere modificar su tasa de mutación y tipo de recombinación en cada caso que se va presentando para optimizar su eficiencia y en algunos casos es simplemente imposible ajustarlo (como se demostró en los capítulos anteriores).

En este capítulo se usa el conocimiento adquirido en los capítulos previos para crear un nuevo algoritmo, el Rank GA [ 13 ][ 14 ] o “Algoritmo Genético dependiente del Ranqueo”. En este algoritmo, los operadores genéticos actúan en función del ranqueo que tiene cada individuo en una lista en la que se ordenan por su aptitud (de aquí el nombre Rank GA). Se hace énfasis en la justificación teórica del algoritmo en función de los grados de libertad efectivos y en el valor teórico del principio subyacente en su forma de operar. Se muestran las ventajas del

Rank GA sobre un “meta” SGA (un SGA cuyos parámetros son siempre los óptimos para cada caso), sobre todo en cuanto a robustez ante diferentes situaciones lo cual lo hace excelente en los problemas de “caja negra” o variantes con el tiempo.

### 3.1 El Rank GA

Este algoritmo surge a partir del análisis hecho en los capítulos anteriores del SGA bajo diferentes situaciones en donde se identifican sus debilidades principales. Quizás la más importante sea la incapacidad para ajustarse a diferentes situaciones. Se han hecho algunas variantes del SGA que tratan de resolver esto. Una de ellas es hacer que la tasa de mutación cambie con el tiempo [ 29 ] iniciando con una tasa alta para hacer una exploración global y, gradualmente, reduciéndola para hacer una exploración más local. Pero esta solución asume que ya se ha logrado algún avance en la evolución después de algún número “razonable” de generaciones y la exploración debe entonces limitarse, lo cual no se sabe a ciencia cierta. Otra solución ha sido el hacer evolucionar la tasa de mutación codificándola en el genoma [ 5 ][ 59 ][ 55 ]. En este caso la idea es que cuando tasas altas dan mejores resultados entonces sean más usadas y viceversa. El problema con esta solución es que, para que la población pueda salir de subóptimos locales, es necesario usar tasas de mutación altas. Para cuando un individuo logra salir del subóptimo local a otro subóptimo mejor, la tasa de mutación sigue siendo alta (mientras se reajusta) y es más probable que este nuevo subóptimo se extinga por efecto de la mutación antes de que se pueda formar una cuasiespecie alrededor de ese subóptimo. Esto retrasa el proceso evolutivo ya que es necesario seguir buscando desde el subóptimo anterior hasta encontrar nuevamente el subóptimo mejor. Lo ideal en este caso sería poder bajar la tasa de mutación a cero en el momento en que se encuentra algún genotipo con mejor aptitud que la aptitud máxima encontrada hasta ese momento. Así se daría tiempo para que sea posible formar una nueva cuasiespecie alrededor de este genotipo y hasta entonces se volvería a incrementar la tasa de mutación. Pero ¿Cuánto se debe incrementar? Lo más razonable sería incrementarla poco a poco para asegurar la supervivencia de la cuasiespecie dominante en ese momento. Pero este proceso cauteloso requiere mucho esfuerzo computacional y tampoco garantiza que el algoritmo logre mantener la cuasiespecie con el mejor individuo encontrado hasta el momento para su explotación. Conservar la cuasiespecie dominante y hacer una exploración adecuada no siempre son objetivos compatibles y, por lo tanto, variar la tasa de mutación no resuelve nada ya que solo se puede cumplir con uno de estos objetivos. Otras técnicas sofisticadas que han sido publicadas [ 4 ][ 20 ] sufren precisamente de eso, de ser sofisticadas y complejas o de no tener un fundamento teórico que las respalde. El Rank GA presentado aquí tiene el valor de que, además de ser altamente eficaz y eficiente, es sumamente simple y su forma de operar está justificada teóricamente mediante el concepto de cuasiespecies y el del efecto de la recombinación estudiados en los capítulos anteriores. Esto hace más fácil no solo su implementación sino el poder entender el principio básico de su funcionamiento.

### 3.1.1 Mutación proporcional al Ranqueo

A partir del hecho de que la tasa de mutación óptima puede ser desde 0 hasta casi 1, como se encontró en los capítulos anteriores, la primera lección por aprender es asignar a la población un rango de tasas de mutación en lugar de una tasa fija. Se asigna entonces a cada individuo una tasa de mutación distinta, así se podrá cubrir un rango de posibles casos en donde esas tasas son ideales. Estas tasas de mutación se asignan a los individuos en función de su aptitud, a los más aptos se les asignan las tasas más bajas para conservarlos y explotarlos, y a los menos aptos las tasas más altas ya que sus genes no resultaron buenos y deben ser mutados con mayor frecuencia para explorar el espacio de búsqueda. La idea fundamental es poder hacer las dos cosas al mismo tiempo, explotar lo bueno y explorar destruyendo lo malo.

El rango de tasas de mutación posibles se define mediante una  $p_{min}$  y una  $p_{max}$ . A partir de lo visto en los capítulos anteriores y para cubrir cualquier eventualidad, se eligió el rango más natural que es con  $p_{min} = 0$  y  $p_{max} = 1$ . Solo en caso de que se sepa que la población está cerca del óptimo es que podría ser mejor usar una  $p_{max}$  más baja pero, en general, no se puede saber esto ya que el paisaje puede o no ser engañoso.

Para asignar a un individuo  $x$  una tasa de mutación  $p_x$  se usa la siguiente formula

$$p_x = p_{min} + (p_{max} - p_{min}) \frac{Rank(x) - 1}{s - 1}, \quad s \geq 1$$

donde la función  $Rank(x)$  indica la posición que ocupa, o el ranqueo de, el individuo  $x$  en la población al ordenarla por el valor de aptitud y  $s$  es el tamaño de la población. Así, la tasa más baja es siempre  $p_{min}$  y la más alta  $p_{max}$  mientras que las demás están uniformemente distribuidas entre estos límites.

Este ranqueo de los individuos hace que la información, obtenida del paisaje de aptitud mediante la población, sea retroalimentada al operador de mutación. Este es el razonamiento Lamarkiano en el que la experiencia adquirida por los individuos se refleja en cambios genéticos que luego son heredados a sus descendientes. Aquellos individuos que en su vida no reciben presión del medio para hacer estos cambios son los que ya están bien adaptados, es decir, los que ya tienen un buen nivel de aptitud ante el medio actual y por lo tanto no cambian mucho y viceversa. Esto implica, además, que existe un grado de comunicación mínimo entre los individuos en el que es posible intercambiar información respecto al grado de “satisfacción” relativo entre ellos para así poder hacer el ordenamiento y saber hasta qué grado es necesario hacer cambios.

Con este esquema se tienen siempre individuos con un nivel de exploración alto pasando por individuos con niveles medios y bajos también. Esto permite hacer exploración tanto global como local, a partir del mejor genotipo encontrado (el cual no muta y solo genera copias de sí mismo al aplicarse selección), permitiendo que la población escape de cualquier óptimo local (aún si éste es muy robusto) conservando siempre la mejor señal que se haya conseguido (aún si ésta es muy frágil ante mutaciones).

En otras palabras, con este operador por ranqueo se modifica la forma de la cuasiespecie alrededor del mejor individuo que se haya encontrado. Cerca de este individuo la cuasiespecie se hace más baja y lejos se hace más alta. Esto significa que habrá menos individuos buscando cerca con un paso de mutación pequeño (pocas mutaciones a la vez) y más explorando lejos con un paso de mutación grande (muchas mutaciones a la vez).

### 3.1.2 **Recombinación con apareamiento de Ranqueo cercano**

En todas aquellas posiciones del genotipo en las que los individuos que se recombinan tienen genes iguales, la recombinación tiene un efecto nulo ya que, en esas posiciones, los hijos no pueden tener genes distintos. Son las diferencias genéticas de los padres las que realmente son recombinadas.

El uso de mutación proporcional al ranqueo puede producir individuos desde los que son iguales al mejor (y buenos) hasta individuos generados aleatoriamente y, por lo tanto, muy diferentes al mejor (normalmente malos). Si se usa el operador de recombinación con emparejamiento aleatorio, sería posible recombinar individuos que son completamente diferentes e incompatibles resultando en un operador demasiado destructivo. Aplicando la misma filosofía que en el caso de la mutación, se diseñó un operador de recombinación basado también en el ranqueo capaz de evitar tal situación. Este operador limita el apareamiento a aquellos individuos cercanos en la lista ordenada por aptitud, es decir, que su ranqueo no es muy diferente entre ellos.

Con este operador, es más probable que los genotipos de los mejores padres sean también similares entre ellos y, por lo tanto, más compatibles en el sentido de que la recombinación entre ellos será menos destructiva ya que tienen pocas diferencias genéticas. En caso de no tener genética muy similar a pesar de estar cerca y en los primeros lugares de aptitud, se deduce que su ranqueo alto se debe a que contienen buenos genes en diferentes posiciones del genoma. Esto hace aun más productiva a la recombinación entre ellos porque sus descendientes podrían resultar con ambas características que los hacen buenos y quizás obtener ambas ventajas.

Para el mejor individuo la recombinación prácticamente no tiene efecto ya que éste tenderá a recombinarse con individuos iguales a él, conservándose el elitismo implícito de la mutación por ranqueo (cuando  $p_{min} = 0$ ). Por otro lado, los individuos generados aleatoriamente (por su alta tasa de mutación) se recombinan con otros igualmente aleatorios resultando también en individuos con genética aleatoria, es decir, que el efecto de la recombinación es aleatorio (como se vio en el capítulo anterior).

La idea de recombinar padres genéticamente similares (y no similares en ranqueo) se ha probado anteriormente por ejemplo en [ 31 ] [ 32 ] en el contexto de la Optimización Multiobjetivo Evolutiva (EMO). En ese trabajo se reporta una convergencia más rápida hacia el llamado “Frente de Pareto” [ 16 ] (la solución óptima) pero también una baja diversidad genética. Para resolver esto fue necesario ajustar los parámetros del algoritmo a cada problema. Con el esquema que se aquí se propone no es necesario ajustar parámetros al pasar de una situación a otra.

Para los experimentos se eligió un emparejamiento con una diferencia máxima de ranqueo del 5% del tamaño de la población (arbitrariamente ya que el método es muy robusto ante cambios en este parámetro y prácticamente cualquier valor bajo es igualmente bueno). Para garantizar el elitismo (conservación del mejor individuo) se forzó al mejor individuo a que nunca se recombine aunque, como ya se vio, lo más probable es que se recombine consigo mismo así que bien podría quitarse este aspecto sin alterar los resultados significativamente. La probabilidad de cruzamiento en todos los demás individuos se fijó en 1.

### **3.1.3 Evaluaciones intermedias**

Ya que los dos operadores descritos arriba tienen efectos muy distintos en la población y que cada uno depende del ranqueo de la población, es necesario hacer este ranqueo antes de aplicarlos. Esto implica que se deberá evaluar a la población con cada ranqueo, es decir, que en una generación completa se evalúa a los individuos dos veces. Para evitar un exceso de llamados a la función de aptitud, a los individuos se les reevalúa solamente cuando hayan sufrido cambios en sus genes. Al reevaluar después de aplicar cada operador se garantiza que el progreso dado por un operador no sea destruido por el otro.

En la naturaleza se puede ver también un fenómeno similar a esta reevaluación entre la mutación y la recombinación si se concibe a la mutación como un operador que actúa durante un período largo, que va desde que nace un individuo hasta que madura lo suficiente como para reproducirse. En este lapso el individuo puede darse cuenta de la necesidad de llevar a cabo cambios genéticos (mutaciones) para lograr una mejor adaptación al medio los cuales serán transmitidos a sus descendientes (este es el pensamiento Lamarkiano). Esto implica una evaluación del estado del individuo durante la aplicación de la mutación para aplicarla en función de este estado. Por otro lado, al momento de elegir una pareja para tener descendencia hay una nueva evaluación del resultado de la mutación prefiriéndose a aquellos que hayan logrado los mejores resultados de adaptación para así buscar la unión de diferentes soluciones buenas en una sola que tenga las (buenas) propiedades de ambas.

## **3.2 Resultados**

En esta sección se muestran las curvas características de búsqueda (definidas en el Apéndice 2) obtenidas a partir de 100 repeticiones de corridas del Rank GA bajo diferentes casos representativos de situaciones reales posibles.

También se muestran curvas para un “meta” algoritmo genético simple (SGA) el cual es un algoritmo genético simple corrido con muchas tasas de mutación diferentes en un rango tal que contenga a la tasa de mutación óptima para así verificar la ventaja del Rank GA.

La probabilidad de cruzamiento se fijó en 0.9 para todos los casos del SGA ya que este valor es conocido como un estándar. En el Rank GA se fijó, como ya se dijo, en 1 excepto para el mejor individuo.

El tamaño de la población se fijó para ambos algoritmos en  $N+1$  siendo  $N$  la longitud del genoma. La razón para este valor tiene dos vertientes:

- 1) Como la tasa de mutación varía desde 0 hasta 1, este tamaño de población de  $N+1$  hace que el mejor individuo tenga tasa cero y que el segundo mejor tenga una tasa  $p = 1/N$ , es decir, que mutará un bit en cada generación mientras que el tercer mejor individuo tendría una  $p = 2/N$  mutando 2 bits cada generación y así sucesivamente hasta llegar al peor individuo quien mutaría todos sus bits. Esto hace al segundo mejor individuo un “hill climber” o escalador local, ya que, normalmente, será una copia del mejor individuo y con una mutación hará búsqueda local alrededor de éste. Si se usara una población más grande, habría más de un escalador local propiciándose la reevaluación de algunas secuencias genéticas ya probadas anteriormente, desperdiándose tiempo computacional.
- 2) En el SGA, la probabilidad de supervivencia de un genotipo que ha sido recién encontrado y que tiene una mejor aptitud que sus vecinos, es alta para este tamaño de población para un rango razonable de tasas de mutación (ver sección 1.7.1.2 y la **Figura 1-49** en la página **84**). De usarse poblaciones más grandes, esta tasa de supervivencia no crecería mucho.

Los resultados obtenidos no difieren mucho cuando se usan poblaciones más grandes y, de hecho, favorecen aún más al Rank GA, en general, así que no se incluyen en este trabajo.

### 3.2.1 **Concatenación de $N$ bloques de aguja en un pajar de 1 bit**

En este paisaje la aptitud es la suma de las aptitudes en cada bloque. Por cada no aguja se suma 1 y por cada aguja se suma  $G$ . También se le conoce como “contando unos” ya la aguja en cada bloque (de un bit) es 1. Este paisaje de aptitud es el más fácil de todos ya que todo el paisaje orienta a la población hacia el óptimo global y, tanto la mutación como la recombinación, son capaces por sí solas de llevar a la población a este óptimo.

En la **Figura 3-1** se muestra que la eficiencia del RGA (marcado con  $\times$ ) en comparación con el mejor caso del SGA (rombos) es prácticamente la misma. Es claro que en aproximadamente la mitad de las repeticiones el RGA es más eficiente que el SGA y en la otra mitad es al revés. La ventaja en cada caso es la misma pero invertida así que se puede considerar a ambos igualmente eficientes. Nótese que se está comparando con el mejor caso del SGA y que, comparando con cualquier otro caso, el RGA es mejor en más ocasiones que en las que es peor.



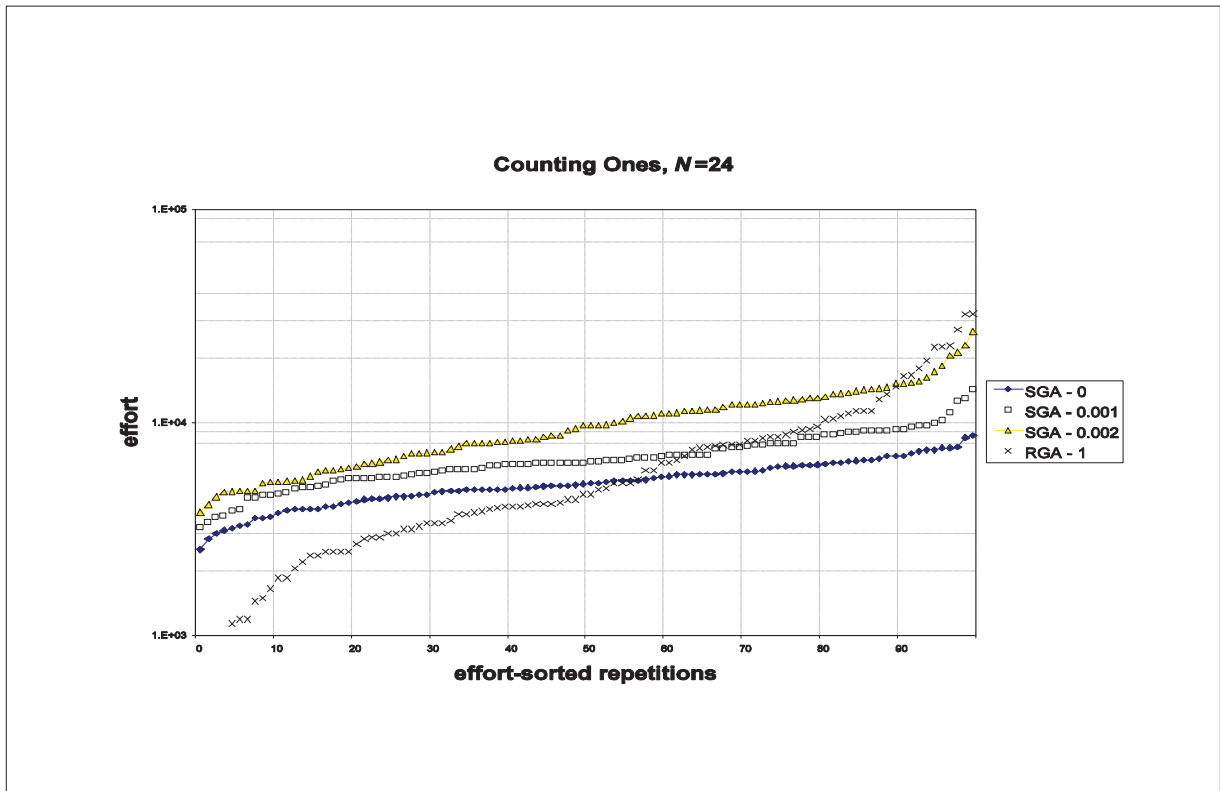


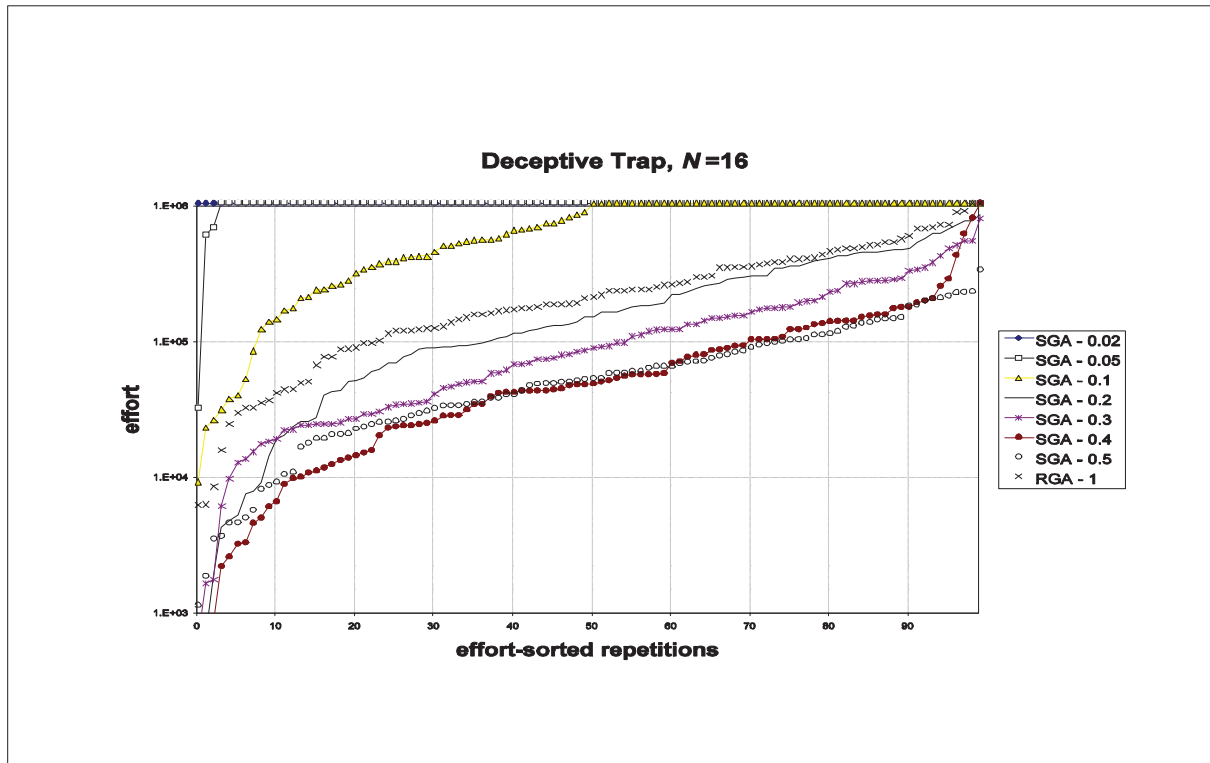
Figura 3-1 Curvas características de búsqueda del RGA y del “meta” SGA en el paisaje de concatenación de 24 bloques de aguja en un pajar de 1 bit.

### 3.2.2 Trampa engañosa

Este paisaje es el mismo que el anterior con la única diferencia de que a la secuencia formada únicamente con ceros se le asigna un nivel de aptitud mayor al de todas las demás convirtiéndola en el óptimo global. En este caso, todo el paisaje es una señal que lleva a la población en dirección contraria al óptimo.

En la **Figura 3-2** se nota que el mejor caso del SGA es igual a la búsqueda aleatoria, lo cual es de esperarse dado que en el paisaje no hay información útil para llegar al óptimo, es decir, que no hay correlación entre el paisaje y el óptimo. El RGA tiene una eficiencia un poco menor que la de la búsqueda aleatoria. Hay que resaltar que, normalmente, el SGA usa tasas de mutación bajas, para las cuales, la búsqueda en este paisaje es sumamente ineficiente y, por consiguiente, su tasa de mutación debe ser modificada en este caso. En contraste, el RGA sigue siendo el mismo que en el paisaje anterior. Si se consideran ambos paisajes de aptitud con el mismo SGA, se verá que solamente puede ser eficiente en uno de los dos o medianamente eficiente en los dos. El RGA puede ser óptimo en el paisaje “contando unos” y al mismo tiempo puede ser casi óptimo en la trampa engañosa lo que demuestra su robustez ante cambios drásticos en el paisaje de aptitud.





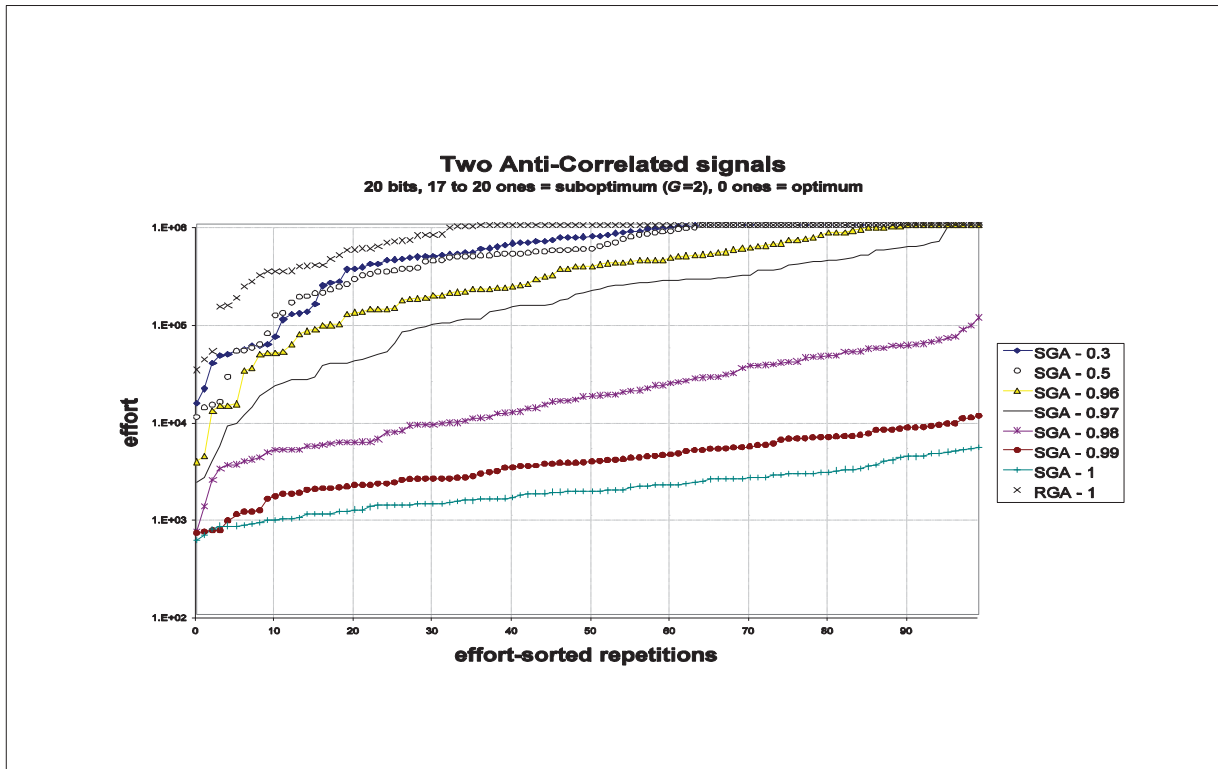
**Figura 3-2** Curvas características de búsqueda del RGA y de varios casos del SGA con diferentes tasas de mutación (incluyendo la mejor posible) en el paisaje de *trampa* engañosa de 16 bits.

### 3.2.3 Dos señales anticorrelacionadas

En este caso se tiene una señal subóptima ancha (formada por varios genotipos o degenerada) en el extremo opuesto de la señal del óptimo, que es angosta (un solo genotipo) y el resto del paisaje es plano. Este paisaje también es engañoso ya que es más probable que en la búsqueda se encuentre primero la señal anticorrelacionada con el óptimo (por ser más ancha) que al óptimo. Sin embargo, en este caso, el SGA tiene muy buena eficiencia con tasas de mutación cercanas a 1. Esta diferencia puede explicarse de la siguiente manera. En el momento en que la población es atraída hacia el subóptimo, la mutación (cercana a 1) la lleva cerca del óptimo, entonces, en el caso de la trampa, la población es repelida del óptimo para luego mutar hacia el subóptimo quedando igualmente alejada de éste de lo que estaba dos generaciones atrás. En cambio, para el paisaje con dos señales anticorrelacionadas, la población no es repelida del óptimo y, por lo tanto, cada vez se acerca más al subóptimo y, alternadamente, al óptimo. En otros paisajes similares pasaría lo mismo siempre que la fuerza de repulsión desde el óptimo fuera menor a la fuerza de atracción al subóptimo.

Véase en la **Figura 3-3** que el RGA (marcado con  $\times$ ) tiene una eficiencia un poco menor a la búsqueda aleatoria (círculos blancos en la parte superior). El caso típico del SGA (tasas bajas de mutación) no se grafica simplemente porque es demasiado ineficiente y su curva quedaría muy arriba. Nuevamente se ve que el SGA, aunque puede ser muy bueno, debe usar una tasa

de mutación muy diferente, en este caso, para ser eficiente mientras que el RGA sigue siendo igual de eficiente sin hacer cambios al algoritmo. Cabe hacer notar que las tasas de mutación cercanas a 1 son buenas, en el SGA, solamente cuando se tienen paisajes con la característica antes descrita de repeler a la población cerca del óptimo con menos intensidad que con la que se atrae a la población hacia la antípoda u opuesta del óptimo. En la práctica, este tipo de paisajes, en los que hay un subóptimo robusto situado en el extremo opuesto al óptimo global, no es común y, por lo tanto, este resultado tiene valor teórico únicamente.



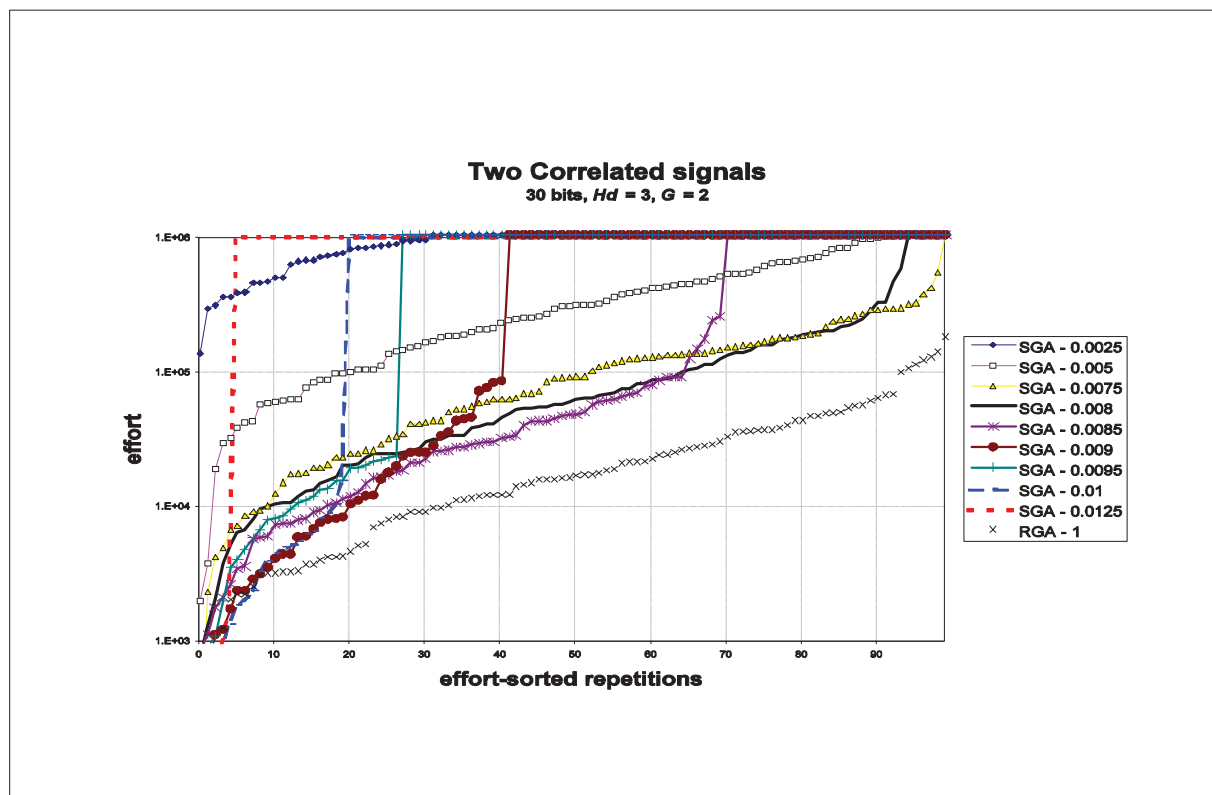
**Figura 3-3** Curvas características de búsqueda del RGA y de varios casos del SGA con diferentes tasas de mutación (incluyendo la mejor posible) en el paisaje con *dos señales anticorrelacionadas*.  $N = 20$  bits.

### 3.2.4 Dos señales correlacionadas

En este caso, como en el anterior, se tienen dos señales pero, ahora, la señal del subóptimo está correlacionada con la del óptimo, es decir, que está situada a distancias de Hamming cortas del óptimo lo que significa que los genotipos de ambas señales tienen más coincidencias que diferencias genéticas. Este paisaje simula casos en donde la población evoluciona de manera puntuada, es decir, que la cuasiespecie dominante “salta” desde una señal subóptima a otra (cuando la segunda es mejor). Para que sea posible dar estos saltos de forma ordenada, es necesario conservar la cuasiespecie presente pero al mismo tiempo explorar el espacio para encontrar una señal mejor y, al encontrarla, se debe poder crear una nueva cuasiespecie alrededor de ésta. En los experimentos se inicializó la población de manera que esté concentrada toda en el subóptimo para medir cuanto tarda el algoritmo (el Rank GA y el

“meta” SGA) en encontrar el óptimo desde allí.

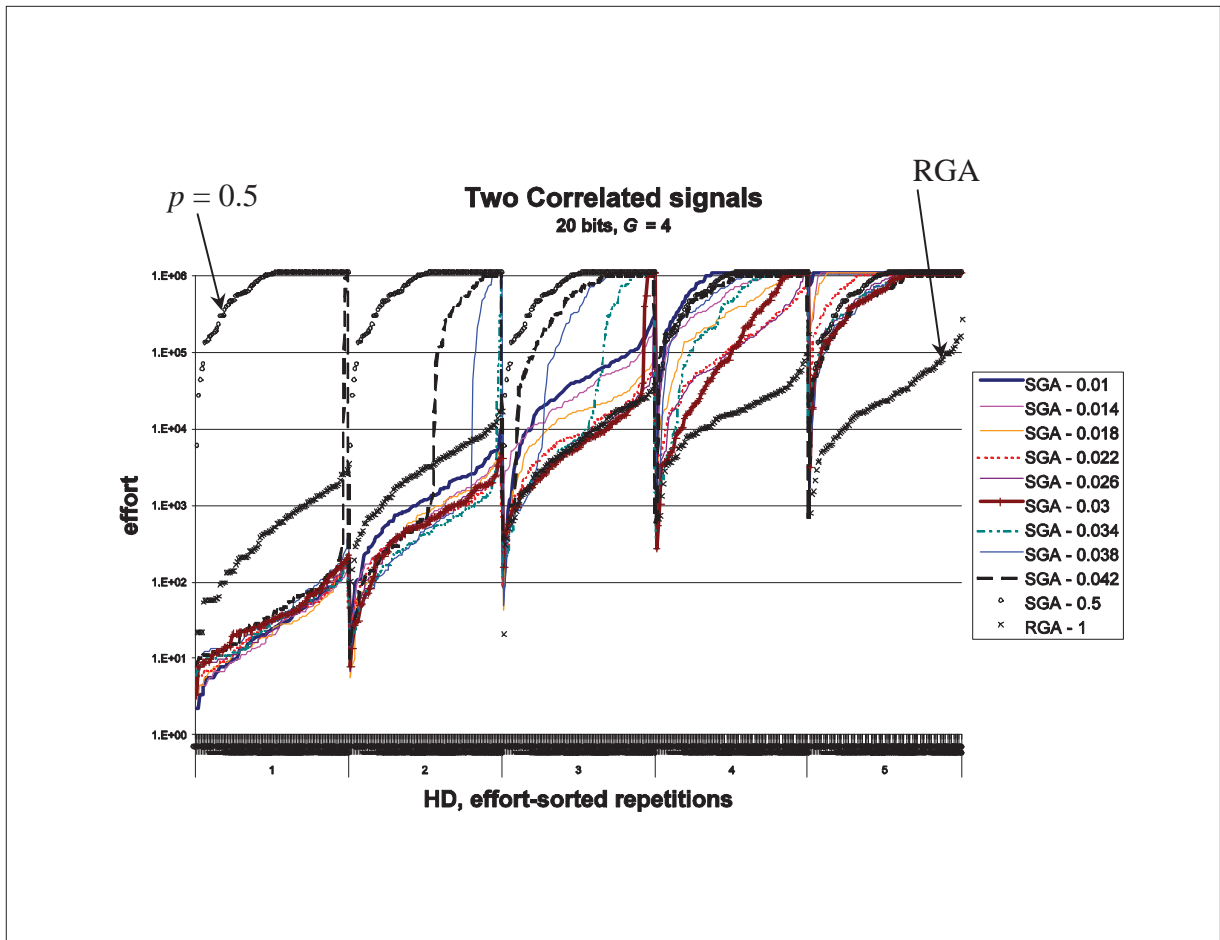
En la **Figura 3-4** se muestra el caso cuando el óptimo está a una distancia de Hamming de 3 del subóptimo. El subóptimo es un solo genotipo con  $G = 2$ , el óptimo también es un solo genotipo pero con  $G > 2$ . Se ve claramente que el RGA es superior a todos los casos del SGA a cualquier tasa de mutación. Esto se debe a que las tasas bajas (curvas siempre altas como la de rombos oscuros) hacen que el SGA tenga poca exploración y las tasas altas (curvas que se doblan hacia arriba casi verticalmente) extinguen, en ocasiones, la señal del subóptimo haciendo que la población se esparza por todo el espacio de búsqueda y que la búsqueda se aleje del óptimo. El RGA nunca pierde la señal presente del subóptimo pero al mismo tiempo puede buscar alrededor de él. Como la búsqueda es más intensa cerca del subóptimo, es más fácil hallar el óptimo.



**Figura 3-4** Curvas características de búsqueda del RGA y de varios casos del SGA con diferentes tasas de mutación (incluyendo la mejor posible) en el paisaje con *dos señales correlacionadas* a una distancia de Hamming de 3.  $N = 30$  bits,  $G = 2$ .

Para ver cómo varía este desempeño como función de la distancia de Hamming entre el óptimo y el subóptimo, se tiene la **Figura 3-5** donde se muestran los resultados para distancias de Hamming desde 1 hasta 5 respectivamente y el subóptimo es una señal con  $G = 4$ . Véase cómo para una distancia de 1 el RGA es un poco menos eficiente que el SGA con tasas bajas, sin embargo, este es el caso más fácil y de hecho el RGA tiene buen desempeño mientras que el SGA depende de que tenga una tasa de mutación  $p$  baja. Al ir aumentando la distancia, el SGA deja de ser tan eficiente hasta el punto de ser, para toda  $p$ , peor que búsqueda aleatoria.

El RGA, aunque sí empeora al aumentar la distancia, es siempre mejor que búsqueda aleatoria. Esto demuestra que el RGA es capaz de seguir la pista de todas las señales correlacionadas presentes en el paisaje.



**Figura 3-5** Curvas características de búsqueda del RGA y de varios casos del SGA con diferentes tasas de mutación (incluyendo la mejor posible) en el paisaje con *dos señales correlacionadas* a varias distancias de Hamming.  $N = 20$  bits.

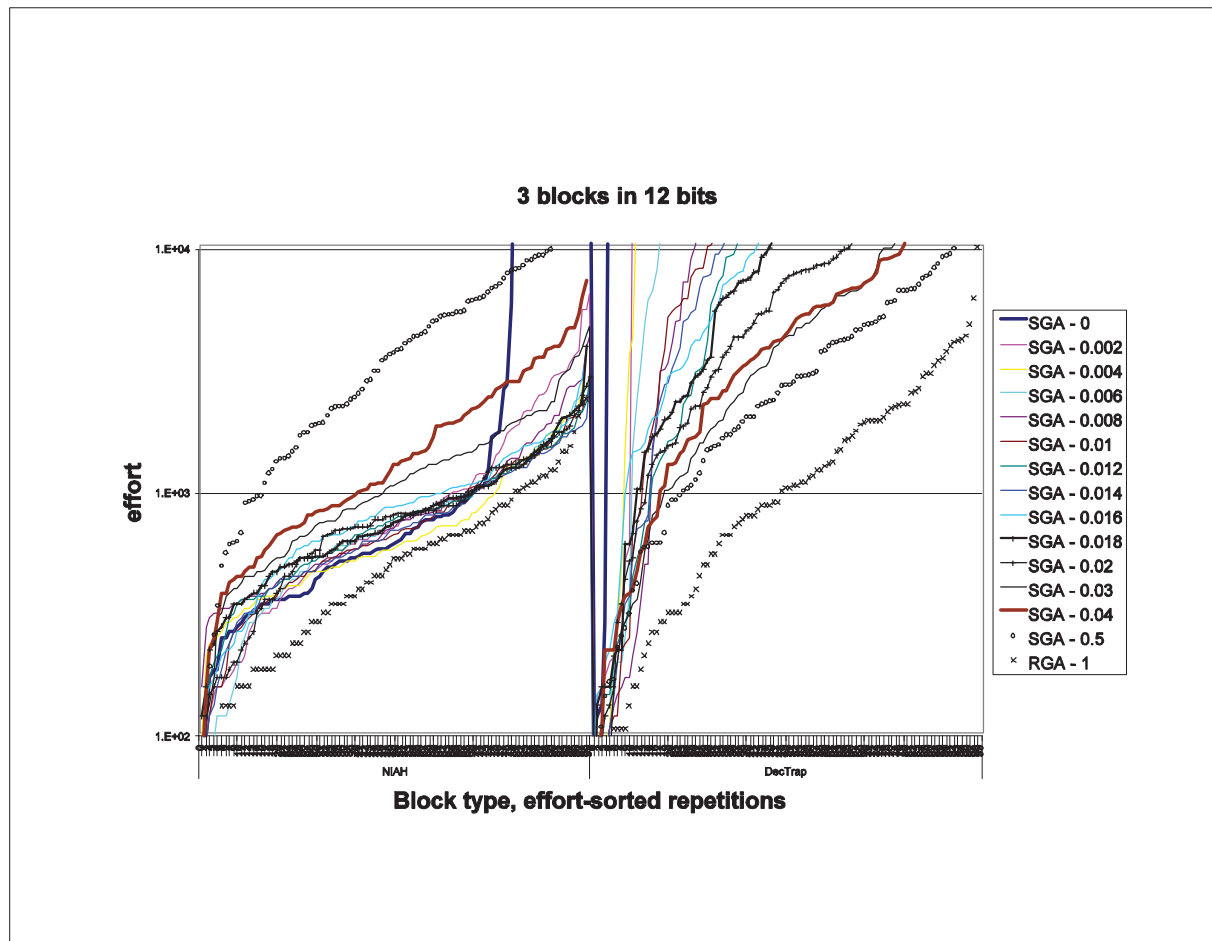
### 3.2.5 Bloques concatenados

La aptitud en estos paisajes es la suma de las aptitudes de cada bloque. Para obtener el óptimo global, es necesario que se consiga el óptimo de cada bloque. Al obtener el óptimo de uno de estos bloques, es necesario que éste se conserve en la población y, al mismo tiempo, seguir explorando el espacio hasta encontrar otro para, luego, juntarlos mediante recombinación y obtener un genotipo con ambos óptimos de bloque y así sucesivamente hasta obtener el óptimo global.

Cuando el paisaje en cada uno de los bloques es de aguja en un pajar, (véase **Figura 3-6** izquierda) se tiene que tanto el Rank GA como el “meta” SGA son mejores que la búsqueda

aleatoria. Sin embargo el RGA (marcado con ×) es mejor que cualquier SGA. La razón de esta ventaja es que el RGA, por ser elitista, nunca pierde lo que ya se ha conseguido, mientras que en el SGA siempre hay una probabilidad de perderlo y, cuando esto pasa, debe buscar más para recuperarlo.

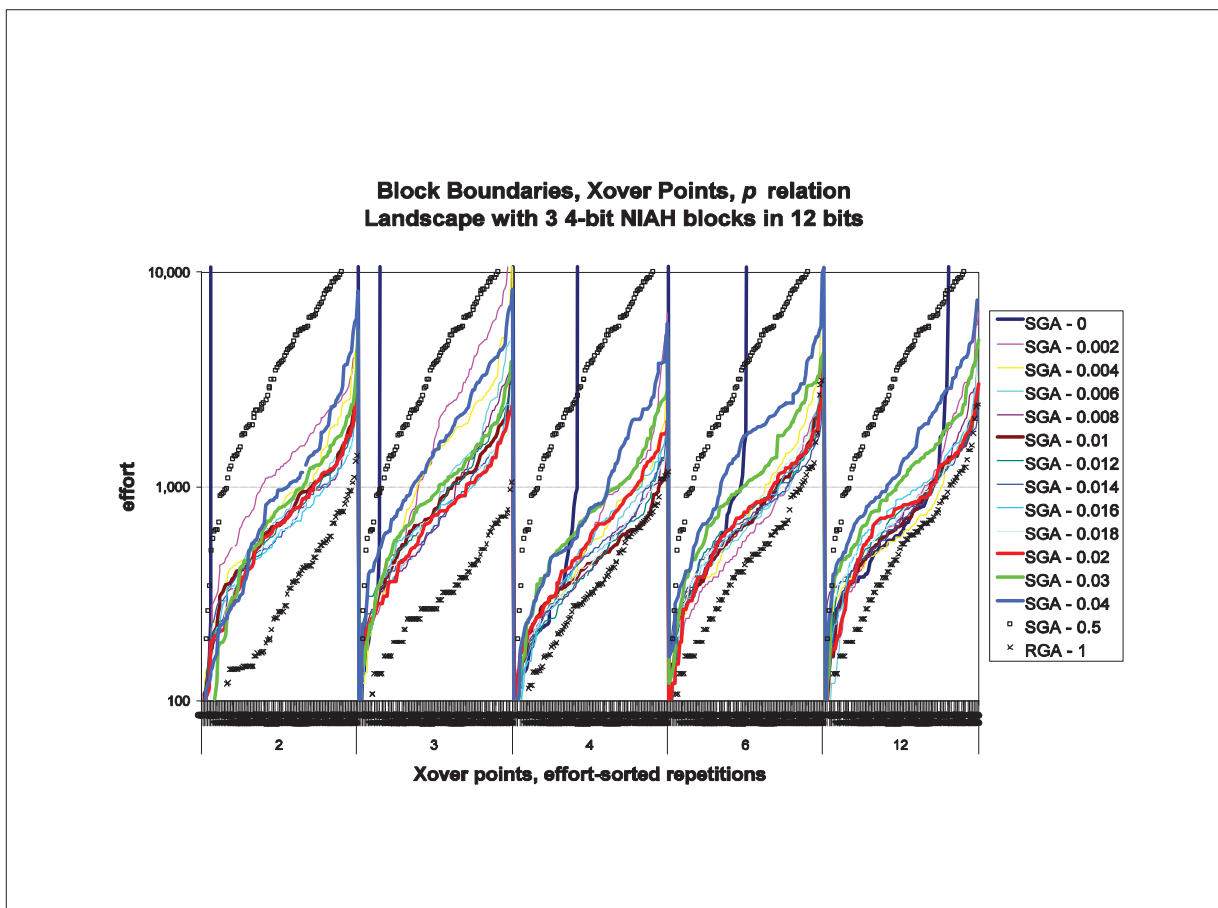
Si los bloques tienen un paisaje engañoso (véase **Figura 3-6** derecha), la situación cambia radicalmente. El SGA es siempre peor que la búsqueda aleatoria mientras que el RGA es mejor. Esto se debe a que la selección en cada bloque va en dirección contraria al óptimo de bloque y, por lo tanto, se requieren tasas de mutación mayores en el SGA para contrarrestarla y acercarse al óptimo (de bloque) haciéndose más destructivo con los bloques ya encontrados. El RGA puede explorar con tasas altas sin perder el óptimo de bloque presente y, cuando un individuo encuentra un óptimo de bloque, es más probable que éste se recombine con otros individuos con algún óptimo de bloque, por lo que se incrementa la probabilidad de juntar ambos óptimos de bloque en un individuo.



**Figura 3-6** Curvas características de búsqueda del RGA y de varios casos del SGA con diferentes tasas de mutación (incluyendo la mejor posible) en el paisaje con tres bloques concatenados de a) aguja en un pajar (izquierda) y b) trampa engañosa, cada bloque de 4 bits.  $N = 12$  bits.

Es evidente aquí que, en el Rank GA, tanto la mutación/selección como la recombinación están optimizadas. La cuasiespecie tiene la forma más adecuada para explorar el espacio alrededor del mejor genotipo encontrado hasta el momento. Los individuos más aptos llevan a cabo una “caminata” aleatoria con pasos pequeños (en el espacio de los genotipos) y los menos aptos lo hacen con pasos grandes lo que permite que la cuasiespecie escape de cualquier subóptimo local. La recombinación se hace entre individuos que han encontrado soluciones parciales igualmente (o casi igualmente) buenas incrementándose así la probabilidad de protegerlas si son muy parecidas y, en caso de ser diferentes, de “sumarlas” en un solo individuo. Todo esto implica una exploración/explotación muy razonable del paisaje de aptitud.

### 3.2.5.1 Bloques concatenados y puntos de cruzamiento limitados

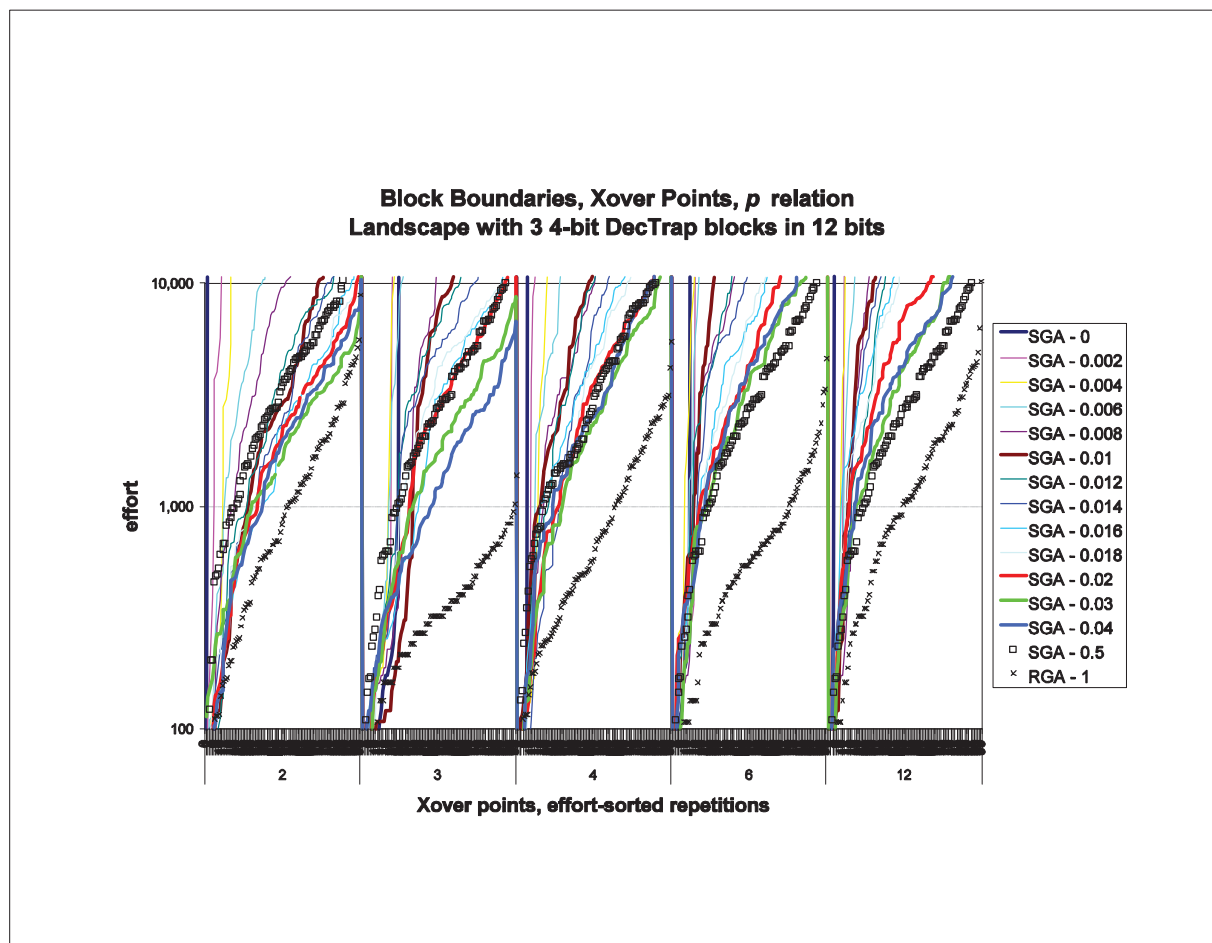


**Figura 3-7** Curvas características de búsqueda del RGA y de varios casos del SGA con diferentes tasas de mutación (incluyendo la mejor posible) en el paisaje con tres bloques concatenados de *aguja en un pajar*, cada bloque de 4 bits, con un número de puntos de cruzamiento elegibles limitado a 2, 3, 4, 6 y 12 respectivamente.  $N = 12$  bits.

En el capítulo anterior se vio que, si se limita el conjunto de puntos de cruzamiento posibles para la recombinación a aquellos puntos en donde hay una frontera en los bloques del paisaje,

la recombinación ayuda a encontrar el óptimo ya que se respeta la integridad de todo óptimo de bloque. Por otro lado, al limitar estos puntos, se tiene también un menor nivel de exploración el cual debe ser compensado con tasas de mutación óptimas más altas. En el Rank GA también se da esta preferencia por limitar los puntos de cruzamiento a las fronteras de bloque. Sin embargo, no es necesario modificar la mutación ya que usa tanto tasas bajas como altas al mismo tiempo.

Para demostrar esto, en los experimentos de esta sección se limitó el número de puntos elegibles como puntos de cruzamiento (en 2 puntos) a 2, 3, 4, 6 y 12 puntos distribuidos uniformemente en un genoma de 12 bits. Por ejemplo, para el caso con 3 puntos, éstos serían en las posiciones 0, 4, y 8 del genoma los cuales coinciden con las fronteras de los 3 bloques en los que se dividió el genoma.



**Figura 3-8** Curvas características de búsqueda del RGA y de varios casos del SGA con diferentes tasas de mutación (incluyendo la mejor posible) en el paisaje con tres bloques concatenados de *Trampa engañosa*, cada bloque de 4 bits, con un número de puntos de cruzamiento elegibles limitado a 2, 3, 4, 6 y 12 respectivamente.  $N = 12$  bits.

Cuando los bloques son de aguja en un pajar (ver **Figura 3-7**), el RGA supera en todos los



casos al meta-SGA, aunque ambos son mejores que búsqueda aleatoria (cuadros). Se nota, sin embargo, que esta ventaja es mayor cuando se limita a 3 puntos posibles de recombinación, en donde el RGA tiene la mayor ventaja.

Si los bloques son engañosos (**Figura 3-8**), la ventaja del Rank GA es aun más grande. El mejor desempeño de ambos algoritmos es cuando se respeta la estructura modular del paisaje con 3 puntos de cruzamiento. En este caso el RGA es muy superior al meta-SGA. Este es el caso donde mejor se demuestran las bondades del RGA. Nótese que el desempeño del SGA es siempre cercano al de la búsqueda aleatoria (cuadros). El RGA es siempre mejor que búsqueda aleatoria debido a que, en el paisaje de aptitud, el óptimo global está correlacionado con los óptimos de bloque, hecho que el RGA aprovecha bien para llegar a este óptimo global, aún cuando la recombinación no respeta las fronteras de bloque.

### 3.3 Conclusiones

El Rank GA y los operadores dependientes del ranqueo son una alternativa interesante al SGA. Su origen proviene del hecho de que tanto la mutación como la recombinación pueden tener efectos constructivos y destructivos de los genotipos más aptos dependiendo de las circunstancias presentes. La mutación y la recombinación son entonces modificadas de manera que sus propiedades permitan lograr una búsqueda más dirigida hacia el aprovechamiento razonable de la información adquirida al ir explorando el paisaje de aptitud.

La mutación por ranqueo reutiliza la genética del mejor individuo que se haya obtenido y hace una exploración centrada en este genotipo, es decir, la cuasiespecie está siempre centrada en este genotipo. Los mejores individuos sufrirán pocas modificaciones aprovechándose así casi todos sus genes lo que implica una búsqueda local mientras que los peores sufren muchos cambios lo que implica una búsqueda global. Esto modifica la forma de la cuasiespecie haciéndola más alta (más individuos) globalmente y más baja localmente (menos individuos). De este modo se cubre una amplia gama de posibles situaciones en las que la búsqueda óptima va desde la local hasta la global impidiéndose siempre la pérdida del mejor genotipo y permitiéndose siempre la búsqueda fuera de toda cuenca de atracción.

La recombinación por ranqueo protege a los mejores genotipos ya que entre éstos hay muchas coincidencias genéticas haciendo que muchas máscaras de recombinación resulten no destructivas o menos destructivas. También, en caso de que el genotipo óptimo esté formado por diferentes bloques constructores, se incrementa la probabilidad de unir estos bloques en un individuo cuando los padres en una recombinación contienen bloques diferentes.

En las aplicaciones prácticas, se diseñan los paisajes de aptitud calificando las soluciones a través de una función de aptitud. En casi todos los casos, se genera así un paisaje de aptitud con varios óptimos locales (o señales), resultando unos más robustos y otros menos. Estos óptimos locales, generalmente, tienen alguna correlación con el óptimo global, sin embargo, puede ser que sean demasiado robustos lo que hace que la búsqueda se concentre demasiado en ellos y se hace imposible hallar el óptimo global. Normalmente, esta robustez no se toma en



cuenta en el diseño de la función de aptitud y, por consiguiente, las buenas soluciones pueden ser frágiles y los subóptimos locales robustos. Por lo tanto es necesario un algoritmo que use la información del paisaje sin dejarse llevar por la robustez de las soluciones. En el SGA, las señales muy robustas pueden llegar a desviar la atención de la búsqueda aún cuando no son muy aptas. El RGA es capaz de seguir en todo momento las correlaciones en las señales presentes en el paisaje de aptitud a partir de la mejor solución conocida pudiendo escapar de cualquier óptimo local.

El comportamiento del Rank GA tiene analogías interesantes en la naturaleza. La población se comporta como una sociedad en la que cada individuo está obsesionado con ser más exitoso que cualquier otro. En este sentido, el grado de libertad efectivo es el grado de éxito relativo entre los individuos. Para incrementar su éxito, un individuo debe saber en qué lugar está en la lista ordenada por “éxito” o aptitud. Así, si es el mejor, no hará modificaciones a su configuración sino que se aferra a su configuración actual ayudando a generar copias de si mismo en la población. Si es exitoso, pero no es el mejor, hará cambios con cautela. Los que están en los últimos lugares saben que es posible mejorar así que intentan cambios arriesgados ya que tienen poco que perder. Esta filosofía evolutiva de poblaciones puede, en principio, ser aplicada en otros campos de la computación evolutiva y en los sistemas adaptativos como, por ejemplo, en el entrenamiento y adaptación de redes neuronales. En este caso podría hacerse una red neuronal en la que se sabe qué tan exitosa es cada neurona en función de la importancia (aptitud) que tenga cada una en la salida de la red. Las neuronas más importantes o exitosas deberán sufrir pocos cambios y las menos importantes podrán hacer grandes cambios en busca de hacer un mejor trabajo que las neuronas que ya lo hacen bien. Así la red es suficientemente robusta para hacer el trabajo mientras las condiciones sean invariantes pero lo suficientemente flexible como para conseguir siempre una mejor adaptación al medio.

En general, el concepto de los operadores dependientes del ranqueo, parece ser un buen principio para hacer colaborar a los individuos con la causa de la sociedad. En este capítulo se presentaron estos operadores en el contexto de los algoritmos genéticos con resultados favorables. Sería interesante realizar experimentos en otros ámbitos para corroborar sus propiedades.

## Notas

## *Capítulo 4*      **Mapeos Genotipo- Fenotipo complejos con neutralidad**

En los capítulos anteriores se usaron paisajes de aptitud (mapeos genotipo-fenotipo-aptitud) modelo con rasgos que representan situaciones extremas. Sin embargo, los mapeos genotipo-fenotipo en problemas “reales” suelen ser complejos generándose toda clase de señales y situaciones en un mismo paisaje. Uno de estos rasgos es la neutralidad ante cambios genéticos en el nivel de aptitud de los individuos dando lugar a las llamadas “redes neutrales”. En mapeos genotipo-fenotipo complejos, puede haber varios niveles de neutralidad, por ejemplo: cuando diferentes codificaciones de un gen tienen la misma expresión en el fenotipo, o cuando algunos genes son desactivados y ya no se expresan en el fenotipo quedando sin efecto cualquier mutación en ellos, o cuando fenotipos diferentes tienen el mismo efecto en la aptitud, etc. Otro rasgo muy común es que se tengan múltiples óptimos locales. Estos óptimos locales pueden o no resultar importantes en la evolución de la población dependiendo de su nivel de aptitud, grado de neutralidad y cercanía con el óptimo global. En este capítulo se analiza un sistema genético bajo un mapeo genotipo-fenotipo complejo para explicar su dinámica en términos de los GLE hallados anteriormente en los paisajes de aptitud modelo y ver así la utilidad de haberlos identificado al ser aplicables en problemas más complejos. Este mapeo está inspirado en el concepto de redes regulatorias genéticas en donde hay genes capaces de activar y desactivar otros genes, de manera que puedan o no ser expresados (que influyan) en el fenotipo. Las características de los paisajes modelo probados anteriormente como son: rugosidad (múltiples señales o ruido), decepción (señales engañosas), modularidad y neutralidad en distintos grados, están todas presentes en este mapeo lo que lo hace un buen caso de estudio.

En la primera parte del presente capítulo se estudia la manera en que la neutralidad afecta la evolución en un algoritmo genético. En la segunda parte se hace un análisis desde el punto de vista de la emergencia de un lenguaje algorítmico [ 3 ] en la genética de la población, en donde la preferencia por ciertos genotipos se describe mediante reglas, parecidas a las reglas gramaticales del lenguaje hablado, a diferentes niveles. Después se usa este modelo complejo para comparar empíricamente el desempeño del SGA con el del Rank GA en diferentes instancias del mismo, desde las fáciles de resolver hasta las muy difíciles. De esta manera se puede tener una idea más clara de las limitaciones del SGA y del potencial del Rank GA en paisajes de aptitud complejos de problemas reales los cuales, normalmente, presentan varias dificultades para hallar el óptimo de manera eficiente.

## 4.1 Neutralidad e intrones

En la teoría de la evolución neutral [ 36 ] se establece que las diferencias que hay entre los genomas de las especies existentes son mayoritariamente neutrales desde el punto de vista selectivo. Es decir, que estas diferencias no influyen en la aptitud reproductiva de las especies o de los individuos que las forman. Se afirma que estas diferencias no pueden ser explicadas como producto de la *selección natural* y que la mayor parte del cambio evolutivo se debe a la *deriva genética* en regiones neutrales del espacio de los genomas.

Esto se basa, en parte, en que hay codones (secuencias de 3 nucleótidos) que, siendo diferentes, codifican el mismo aminoácido. Así, algunas mutaciones en los nucleótidos podrían resultar “silenciosas” o “no expresadas”. Por otro lado, se sabe que algunos genes actúan como reguladores en la expresión de otros genes, esto es, que determinan hasta que grado se “usan” otros genes. Estos genes reguladores forman una Red de Regulación Genética que puede desactivar genes convirtiéndolos en “intrones”, de manera que las mutaciones en ellos ya no tienen ningún efecto, es decir, que son neutrales. Otro nivel de neutralidad es cuando dos (o más) fenotipos distintos resultan igualmente buenos en cuanto a aptitud, entonces, si debido a cambios genéticos cambia uno de esos fenotipos y se transforma en el otro, habrá sido un cambio neutral.

Los intrones son material genético ubicado entre los genes que no tiene influencia en el fenotipo. La presencia de los intrones provoca *neutralidad* en el mapeo genotipo-fenotipo ya que, para muchos genotipos distintos se tiene el mismo valor de aptitud selectiva.

La neutralidad tiene la desventaja, en el ámbito de la Computación Evolutiva, de hacer más grande el espacio de búsqueda lo cual incrementa el esfuerzo computacional, pero se ha probado que, en muchas situaciones, tiene la ventaja de permitir mayores oportunidades de evolución. Este es el caso cuando los intrones pueden eventualmente dejar de serlo y convertirse en material genético codificador. También los intrones, aunque no lleguen a codificar, pueden ayudar a que el operador de cruza sea menos destructivo al funcionar como separadores de material genético que no debería ser truncado y de esta manera aumenta la probabilidad de que el material genético se cruce con material genético compatible. Se ha observado otra desventaja y es que, en ocasiones, los intrones pueden producir estancamiento en el proceso evolutivo si existe un número excesivo de ellos.

En el caso de evolución de secuencias genéticas de *longitudes distintas*, los intrones juegan un papel muy importante. Por ejemplo, si las secuencias óptimas son aquellas que contienen a cierta subsecuencia de determinada longitud, las secuencias óptimas de mayor longitud tienen un número mayor de intrones y también una mayor resistencia a las mutaciones ya que es posible que la subsecuencia esté codificada en más de una posición en ellas. Al ser más resistentes a las mutaciones, las secuencias más largas tienden a dominar la población generándose el fenómeno del “*bloat*” o hinchazón, en el que la longitud promedio de los genomas en la población va en aumento.

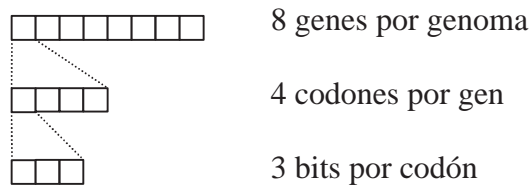
El material genético no codificador (o intrones) tiene una estructura que no es aleatoria. Hay una tendencia a que estas secciones formen material genético de “reserva”. Esta reserva

genética da una cierta protección al individuo contra mutaciones ya que, de ser activada, esta tiende a ser más compatible que el material aleatorio. Esto se debe a que estas secciones no codificadoras alguna vez fueron genes activos y conservan, hasta cierto grado y durante algún tiempo, la forma que tenían cuando estaban activas y eran compatibles.

El grado de neutralidad depende del porcentaje de intrones en las secuencias genéticas, es decir, de la fracción del total de las mutaciones que son “silenciosas” en cada genoma. Este porcentaje puede variar de una región a otra en el espacio de los genomas.

## 4.2 Un mapeo genotipo-fenotipo complejo con neutralidad inspirado en las Redes de Regulación Genéticas

El mapeo genotipo-fenotipo que se usa en este capítulo es el descrito en [ 66 ] en donde cada genoma está formado por 8 genes; cada gen está formado por 4 codones y cada codón por 3 dígitos binarios (96 bits en total).



De los 8 genes, el primero se considera el *gen selector* de genes (ver **Figura 4-1**). Cada codón del gen selector es un apuntador hacia alguno de los genes (pueden apuntar al mismo gen selector y/o a un mismo gen varias veces). En el mismo orden en que son seleccionados/activados, los genes son copiados a una secuencia de 4 genes la cual es traducida codón por codón mediante una tabla de traducción predefinida. Esta secuencia es después interpretada de izquierda a derecha como un árbol escrito en preorden considerando a cada codón (traducido) como un generador o un terminador. En este caso se consideró el siguiente significado para cada símbolo en que puede traducirse un codón.

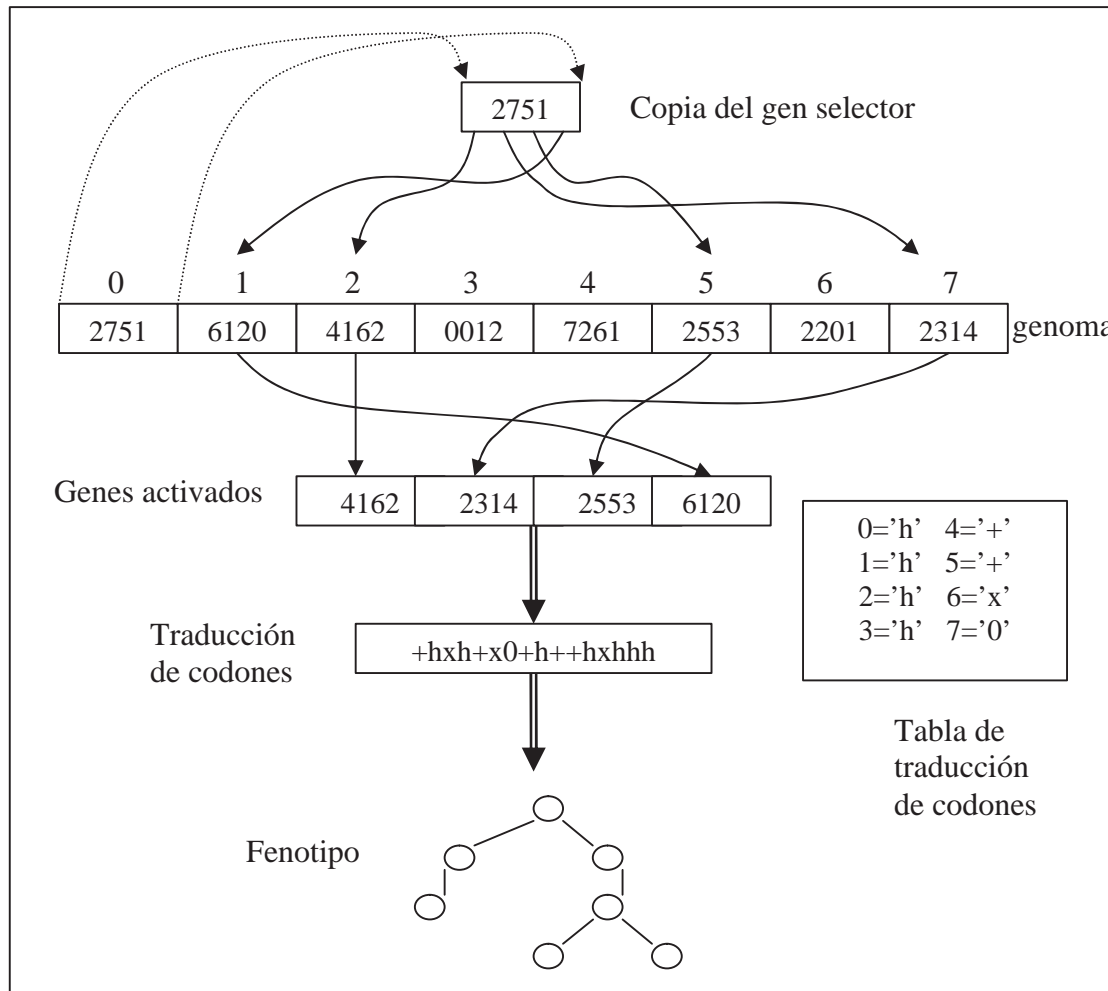
Símbolo	significado	tipo	descripción
h	$\tanh(x)$	generador	tangente hiperbólica de x
+	suma( $x_1, x_2$ )	generador	suma de $x_1$ y $x_2$
x	x	terminador	variable x
0	0	terminador	valor 0

Finalmente el árbol se prueba con un conjunto fijo de valores de la variable x contra la función

$f(x) = 4x$ . Los árboles que mejor aproximen dicha función tendrán un valor de aptitud mayor. Para esto se utilizó la siguiente función de aptitud

$$ft = \frac{100}{1 + error}$$

donde  $error = \sqrt{\sum_i (4x_i - A(x_i))^2}$ , y  $A(x_i)$  es el valor que devuelve el árbol para cada  $x_i$ .



**Figura 4-1 Mapeo Genotipo-Fenotipo usado en este capítulo.**

El paisaje de aptitud asociado con este mapeo genotipo-fenotipo-aptitud presenta neutralidad a diferentes niveles. En primer lugar, el gen selector solamente activa 4 de los 8 genes presentes. Esto significa que por lo menos 3 genes pueden sufrir cambios sin que se afecte el fenotipo y por lo tanto la aptitud. Además, no todos los genes y sus correspondientes codones que son activados son usados en el árbol resultante convirtiéndose en intrones también. Por otra parte, existen varias formas de hacer la misma regresión de la función con diferentes árboles debido a las propiedades matemáticas de ésta. Y, finalmente, hay cierto grado de neutralidad en la

traducción de los codones en símbolos generadores o terminadores ya que varios codones codifican el mismo símbolo.

A pesar de haber tantas posibilidades de neutralidad, el paisaje de aptitud generado es muy rugoso y se forman muchos óptimos locales y globales. Resulta sencillo crear instancias de problemas desde las muy fáciles hasta las muy difíciles de resolver dependiendo de la función de la que se requiera hacer la regresión, lo que da un buen marco para el análisis de sistemas genéticos evolutivos.

### 4.3 Efecto de la neutralidad

Para distinguir el efecto de la neutralidad en este mapeo genotipo-fenotipo, se hizo evolucionar un SGA con una población con los operadores de mutación ( $p = 0.05$ ) y cruzamiento en un punto ( $p_c = 0.9$ ). En la **Figura 4-2** se muestra cuales son los fenotipos preferidos de entre aquellos que representan exactamente a la función  $f(x) = 4x$ . Para obtener esta gráfica se hicieron 31 corridas por 1000 generaciones y en cada generación se ordenó a la población del más apto al menos apto, en caso de aptitudes iguales se ordenaron de manera aleatoria. De cada generación se extrajo al primero de la lista ordenada (el mejor individuo) y se sumó el número de veces que cada fenotipo apareció en este lugar. Nótese que, a pesar de que todos tienen la misma aptitud, hay ciertos fenotipos que dominan mucho sobre otros debido a que unos tienen mayor aptitud efectiva que otros. Se observa que los fenotipos óptimos más frecuentes son aquellos que tienen mayor robustez ante cruzamientos y mutaciones, es decir, los que tienen mayor número de intrones o mayor grado de neutralidad. Véase por ejemplo que el fenotipo ++xx++xx0 es mucho más frecuente que ++++xxx0x ya que el primero puede codificarse por ejemplo con

551\* 0\*\*\* \*\*\*\*\* ++xx \*\*\*\*\*

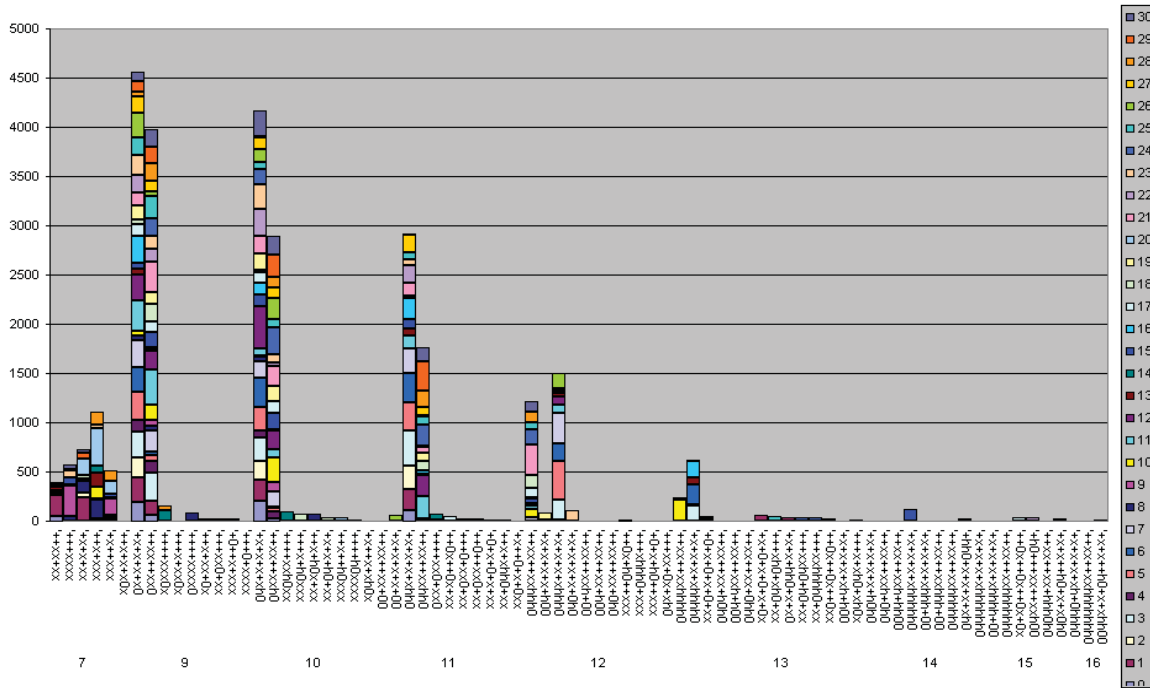
repiéndose la activación de un gen, lo que le da un mayor grado de neutralidad (con 24 codones sin expresar) que el del segundo cuya codificación debe ser algo como

527\* \*\*\*\*\* xxx0 \*\*\*\*\* +++++ x\*\*\*

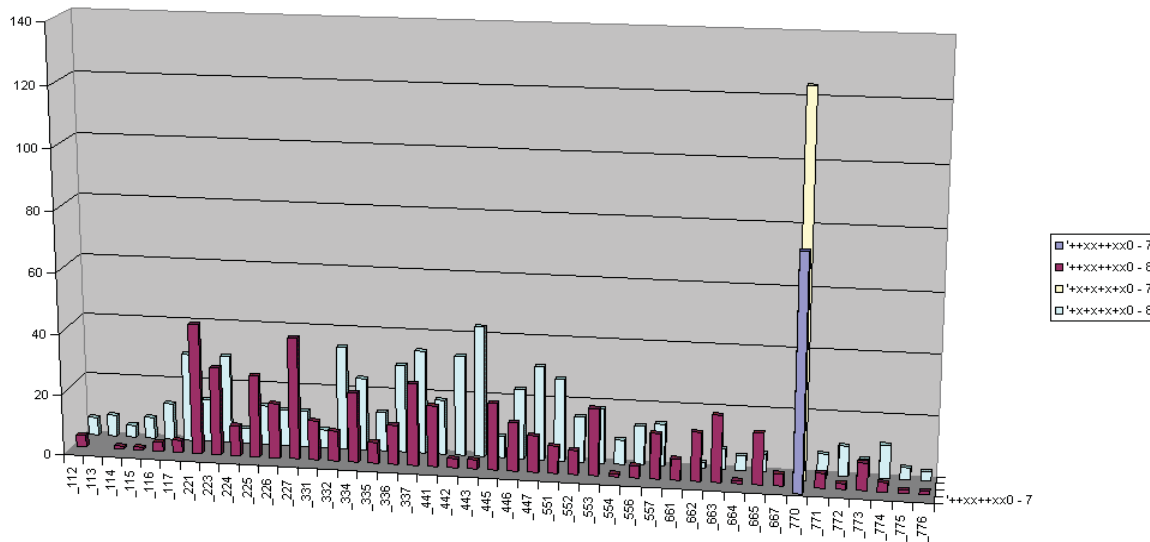
que deja libres (sin expresar) un número menor de codones (20).

En la **Figura 4-3** se muestra la frecuencia con que son seleccionadas las distintas configuraciones del gen selector cuando éste corresponde con el fenotipo +x+x+x+x0 ó ++xx++xx0. Solo se muestran los selectores más frecuentes. Es notable que todos estos activan el mismo gen dos veces (que codifica ++xx o bien +x+x) y luego uno más (que codifica en su primer codón un 0). Véase cómo es que el selector más frecuente es 770\*. La ventaja se explica porque hace uso del mismo gen 2 veces y luego se usa a sí mismo para codificar el cero al final del fenotipo ya que el codón 7, que está al inicio, se traduce en un cero. Al usar solamente 2 genes y a sí mismo realmente se están activando solamente 7 codones: los 3 del selector, y los 4 del gen activado (el siete), en cambio, los demás selectores

activan 8 codones, uno más en otro gen para codificar el cero al final del fenotipo.



**Figura 4-2** Gráfica de la frecuencia con que son seleccionados los diferentes fenotipos óptimos ordenados por su longitud y alfabéticamente.



**Figura 4-3** Gráfica de la frecuencia con que son seleccionadas las diferentes configuraciones del gen selector que codifican ++xx+xx0 y +x+x+x+x0, la notación -7 y -8 indican el número de codones del genoma que se expresan en el fenotipo.



Al haber múltiples soluciones óptimas, es el grado de neutralidad en ellas el que hace a unas más frecuentes que a otras. Esto se debe a que las que tienen más neutralidad ocupan más espacio en el espacio de los genotipos y además, al ser halladas por la población, resultan más difíciles de extinguir ya que resisten más la destrucción por parte de los operadores genéticos, es decir, que son más robustas. En otras palabras, la señal en el paisaje que contiene a estas soluciones tiene una tasa de mutación crítica más alta y además la recombinación es menos destructiva con ellas ya que, al haber más copias de ellas en la población, sus parejas tienden a ser miembros de la misma señal.

#### 4.4 Emergencia de un lenguaje algorítmico

En un mapeo genotipo-fenotipo tan complejo como el descrito antes, pueden distinguirse grados de libertad efectivos distintos a los grados de libertad en mapeos más simples. Así como en un lenguaje hablado hay una cierta estructura que debe respetarse para que lo dicho tenga sentido, en la genética de la población emerge una estructura que puede describirse mediante reglas “gramaticales” a diferentes niveles. Puede asociarse a los codones con las letras, a los genes con las palabras y a los genotipos con las frases de un lenguaje hablado. Durante el proceso evolutivo unas *letras* son preferidas sobre otras por su utilidad para expresar alguna idea, igualmente, ciertas *palabras* tienen preferencia así como ciertas *frases* o secuencias de palabras las cuales cumplen con las “reglas” gramaticales para que sean valiosas. Así, las frases que siguen esas reglas, usando las palabras adecuadas que a su vez están formadas por las letras preferidas tendrán mayor aceptación y por lo tanto mayor probabilidad de supervivencia.

Para ver esto, se hizo una corrida buscando la función  $f(x) = 4x$  con un algoritmo genético simple (SGA) en el cual la recombinación se aplicó con probabilidad 0.9, la tasa de mutación fue 0.01 y la población de 100 individuos. Después de 15 generaciones la población es la que se muestra en el Apéndice 4.

De esta población se puede obtener la frecuencia de cada uno de los alelos o letras que sería el nivel más bajo en la descripción del lenguaje. Puede verse que las frecuencias de los alelos son:

?	h	580
?	+	414
?	x	405
?	0	201

Total 1600

en las que se nota un sesgo selectivo al compararlas con las frecuencias que se tendrían en una población aleatoria, a saber: para h serían 800, para + 400, para x 200 y para 0 200. Hay una marcada preferencia por el uso del símbolo x y un rechazo del símbolo h.

Un nivel más elevado del lenguaje es el de las palabras. Aquí se puede ver que las palabras más usadas (y sus frecuencias) son:

_h+hh	58	_xhhh	7	_+hxx	3	_+h0+	2	_x0hx	2	_0+x+	1	_h0xh	1
_++hx	47	_+0hh	5	_+xxh	3	_+x++	2	_x0xh	2	_0+x0	1	_hh+x	1
_+x+x	45	_0hhh	5	_0h++	3	_+x+0	2	_xh+h	2	_00+0	1	_hh0+	1
_0+xh	41	_h+0h	5	_0hx0	3	_+x00	2	_xhh0	2	_00+x	1	_hhh0	1
_hh0h	40	_hhh+	5	_h0++	3	_+xh0	2	_xhx0	2	_0h0+	1	_hhx0	1
_00hx	35	_hxx0	5	_h0h+	3	_0+h+	2	_xhxx	2	_0hhx	1	_hx0+	1
_0h+x	34	_xhhx	5	_hh0x	3	_00h+	2	_xxh+	2	_0hx+	1	_hx00	1
_0x+x	34	_++hh	4	_hx+0	3	_00hh	2	_+++0	1	_0x+h	1	_hxx+	1
_hhhh	21	_+x+h	4	_hx+x	3	_00x+	2	_++00	1	_0x0h	1	_hxx0	1
_hhhx	15	_0+hh	4	_hx0h	3	_0h+h	2	_++0h	1	_0xh0	1	_hxxx	1
_h+hx	12	_0h0h	4	_hxx+	3	_0hxx	2	_+0++	1	_0xhh	1	_x+++	1
_hx++	11	_h00h	4	_x0hh	3	_0xh+	2	_+0+h	1	_0xx+	1	_x+00	1
_h0+h	10	_h0h0	4	_xhh+	3	_h++0	2	_+0hx	1	_0xxh	1	_x0++	1
_hh+h	10	_h0hh	4	_xx+x	3	_h+h0	2	_+h++	1	_0xxx	1	_xh+0	1
_hxhh	10	_h0x+	4	_xxhh	3	_h0hx	2	_+h+0	1	_h+++	1	_xh00	1
_+hh+	9	_hxx+	4	_++++	2	_hh++	2	_+h0h	1	_h+++	1	_xhx+	1
_+hhh	9	_hxx	4	_++++	2	_hh+0	2	_+h0x	1	_h+00	1	_xx++	1
_hhxh	9	_hxhx	4	_+++h	2	_hh00	2	_+hx0	1	_h+0x	1	_xx+0	1
_+xhh	8	_hxxh	4	_+++x0	2	_x+++	2	_+x0h	1	_h+h+	1	_xx+h	1
_0+++	8	_x+h0	4	_+0+x	2	_x+0h	2	_+xhx	1	_h+x+	1	_xxx+	1
_0hh+	8	_xx0h	4	_+00x	2	_x+h+	2	_0++0	1	_h+x0	1		
_hx+h	8	_++xh	3	_+0h+	2	_x+hx	2	_0++h	1	_h+xh	1		
_+h+h	7	_+hh0	3	_+0x+	2	_x+xh	2	_0++x	1	_h+xx	1		
_0x++	7	_+hhx	3	_+0xh	2	_x0+h	2	_0+0h	1	_h0+0	1		
_x+hh	7	_+hx+	3	_+h+x	2	_x0h+	2	_0+hx	1	_h000	1		

Nótese que hay muchas palabras que inician con 0 que es el terminador que se necesita después de sumar 4 veces x al repetir la palabra +x+x dos veces. También es frecuente que las palabras inicien con h0 o hh0 o hhh0 que se evalúan también como cero pero al ser más largas quedan más expuestas a mutaciones destructivas haciéndolas menos valiosas. Otras combinaciones que también evalúan a cero sufren igualmente por ser más largas y por lo tanto más frágiles. Todas estas diferentes codificaciones de 0 están presentes con mayor frecuencia que otras, aún cuando no se expresen en el fenotipo, funcionando como genes de reserva que hacen más robustas las soluciones en caso de que llegaran a activarse y expresarse debido a alguna mutación en el gen selector. Esto es consecuencia de un rompimiento inducido de la simetría en la aptitud de las soluciones. Es decir, soluciones igualmente aptas no tienen la misma aptitud efectiva.

Si se observa con detenimiento puede verse que algunas palabras son seguidas de unas palabras con mayor frecuencia que de otras. Por ejemplo: la frase \_5526\_7561\_7346\_7706\_7656\_4646\_2270\_2430 es la más frecuente seguida de \_5526\_0136\_0416\_1645\_7454\_4646\_2750\_2410. Queda claro, por ejemplo, que en esta gramática emergente, una regla es que las frases deben iniciarse preferentemente con la palabra \_5526 y que en la 6ª posición debe estar la palabra 4646. Así, en este “lenguaje”, las frases que tienen más utilidad para

intercambiar ideas son las que están bien estructuradas, es decir, que cumplen con esta regla gramatical.

Es claro que dicha estructura no estaba al inicio de la corrida sino que fue emergiendo con las generaciones. Inicialmente son seleccionadas unas letras más que otras, luego unas palabras más que otras y, finalmente, unas frases más que otras.

Nótese que esta estructura gramatical compleja emerge debido a la estructura compleja del mapeo genotipo-fenotipo. De haberse usado un mapeo más simple, habría emergido una gramática más simple también, y viceversa.

Las reglas del lenguaje hablado en la naturaleza pudieron haber surgido de manera similar las que emergen en el modelo. Primero fueron los sonidos simples los que sirvieron como medio de comunicación lo que significó una ventaja reproductiva para aquellos que adquirieron esa habilidad. Luego, al no ser suficientes las distintas señales que podían ser comunicadas con los sonidos disponibles, aparecieron secuencias de sonidos en las que el orden de los sonidos sirvió para distinguir una señal de otra, es decir, las palabras. Sin embargo, algunas de las posibles palabras son preferidas formándose un vocabulario. Más adelante surgen secuencias de palabras con el mismo fin, comunicar una mayor variedad de señales. Pero estas secuencias de palabras o frases surgen con una estructura gramatical para que tengan mayor claridad, es decir, que las frases que cumplen con ciertas reglas son seleccionadas con mayor frecuencia.

Esto hace pensar que este modelo complejo puede servir como herramienta para estudiar algunos aspectos de la evolución del lenguaje desde sus orígenes y, sobre todo, las condiciones necesarias para que haya podido darse.

## 4.5 Búsqueda con operadores dependientes del Ranqueo

Con este paisaje de aptitud complejo, en donde hay muchos óptimos globales (y locales), puede pensarse que sería fácil encontrar por lo menos alguno de ellos con poco esfuerzo. Esto es cierto cuando realmente hay mucha neutralidad en las soluciones ya que el espacio de búsqueda se reduce efectivamente. Sin embargo, es posible reducir el nivel de neutralidad en los óptimos. Una manera es que la función de regresión requiera ser codificada con muchos codones reduciéndose así el número de codones no expresados o neutrales. Otra manera es que esta función no pueda ser expresada matemáticamente de muchas maneras distintas (con los operadores disponibles), por ejemplo, la función  $f(x) = 4x$  puede expresarse como  $3x+x$  o  $x+3x$  o  $2x+2x$  o  $x+x+2x$  o  $x+2x+x$  o  $2x+x+x$  o  $x+x+x+x$  indistintamente, pero la función  $f(x) = x+\tanh(x)$  puede expresarse como  $x+\tanh(x)$  o  $\tanh(x)+x$  solamente. Obviamente también están las opciones en las que se suman ceros en ambas funciones.

Por lo anterior resulta interesante probar el algoritmo presentado en el capítulo anterior (el Rank GA) sobre este modelo en diferentes instancias del mismo, desde las más fáciles hasta las más difíciles. Para lograr instancias de prueba más difíciles se substituyó la traducción del codón 7 en el símbolo '0' por '1' ya que este no es neutral respecto a la suma ni a la tangente

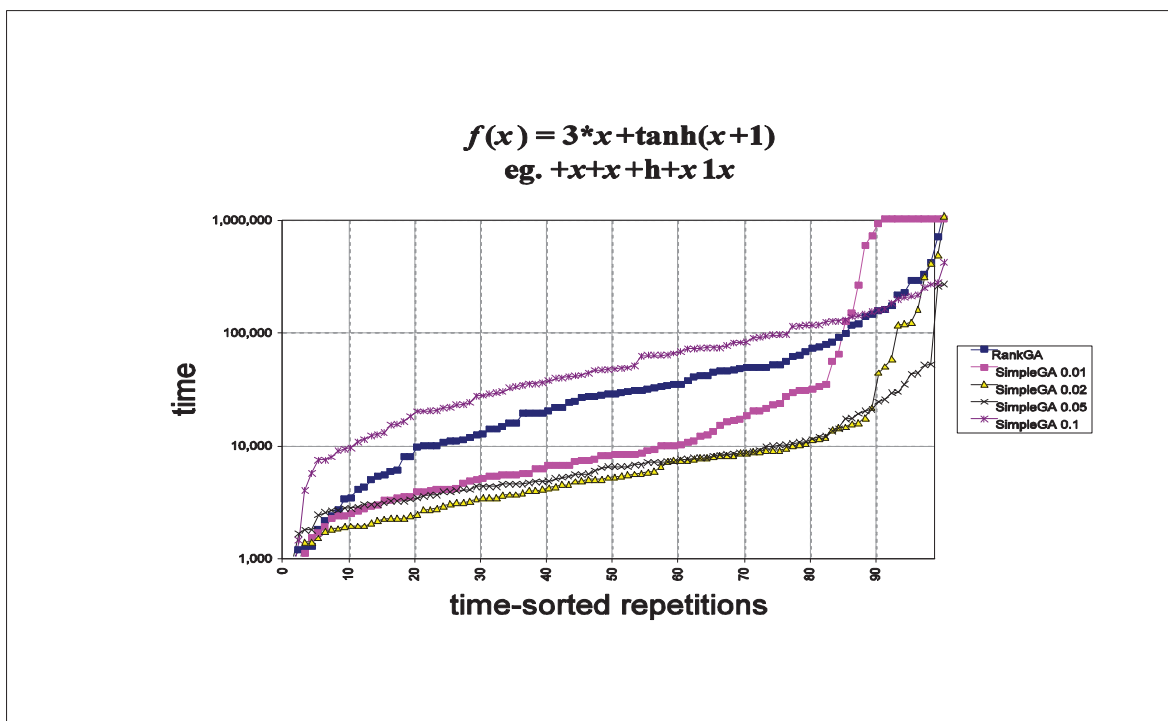
hiperbólica lo que reduce el grado de neutralidad de todas las soluciones.

### 4.5.1 Casos con alta neutralidad

Se usa aquí una instancia del modelo en la que el número de posibles configuraciones óptimas (el grado de neutralidad en las soluciones) es suficientemente grande como para hacer el problema relativamente fácil de resolver (aunque no trivial) tanto para el SGA como para el RGA.

El caso de prueba es aquí el de la regresión de la función  $f(x) = 3x + \tanh(x+1)$ . Esta función tiene varias simetrías que aumentan su grado de neutralidad como por ejemplo, que los términos pueden conmutarse. De hecho, el término  $3x$  realmente debe ser codificado como la suma de  $x$  tres veces y por lo tanto se puede conmutar cada  $x$  con el término  $\tanh(x+1)$ . Los términos en el argumento de  $\tanh$  también pueden conmutarse. Todo esto ayuda al algoritmo ya que, si no se acerca a una forma, se acercará a otra, aumentando la probabilidad de éxito.

En la **Figura 4-4** se muestran los resultados con esta función. Se observa que el SGA ( $p_c = 0.9$ ) puede desempeñarse mejor que el RGA solo cuando la tasa de mutación está dentro del rango de 0.02 a 0.1 y fuera de ese rango es siempre peor. Este rango de tasas de mutación es muy diferente a la heurística común  $1/N$  que en este caso es aproximadamente 0.01. Esto hace difícil la puesta a punto del SGA.



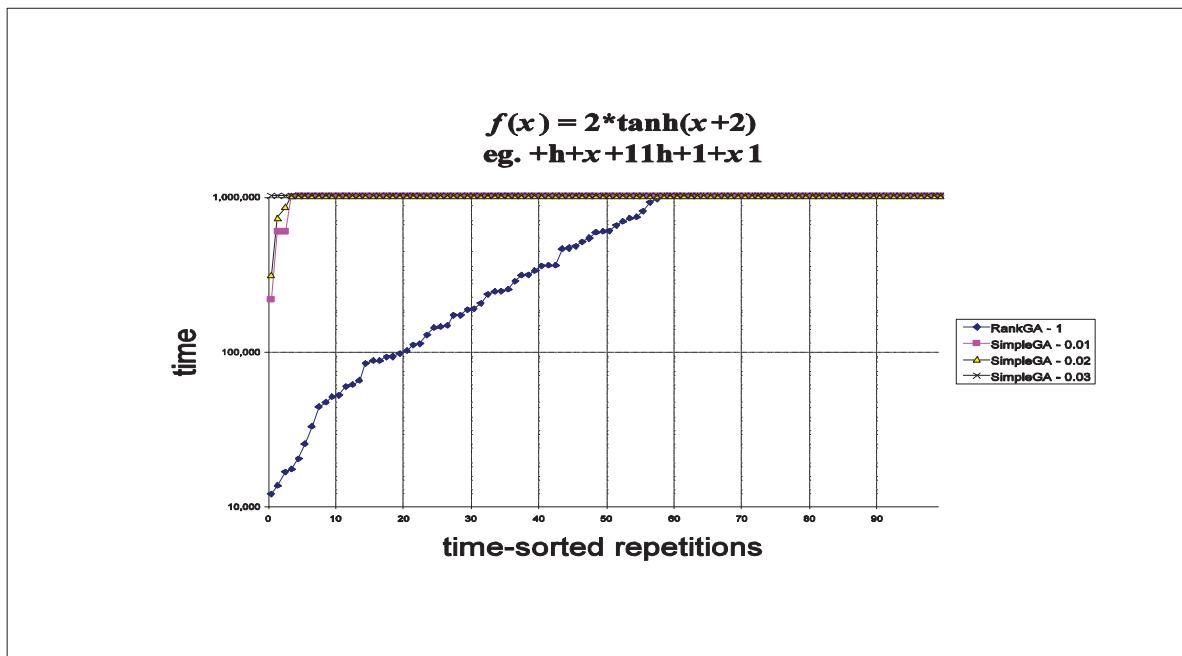
**Figura 4-4** Curvas características de búsqueda del RGA y de varios casos del SGA con diferentes tasas de mutación (incluyendo la mejor posible) en el caso de la regresión de la función  $f(x) = 3x + \tanh(x+1)$ .

Para ajustar el SGA es necesario correrlo hasta encontrar la manera en que puede resolver el problema eficientemente. La tasa ideal de mutación depende del grado de neutralidad en el paisaje (entre muchas otras características de éste) y por lo tanto las heurísticas disponibles, como  $1/N$ , no ayudan mucho a ajustar el parámetro. El resultado es que para resolver un problema de manera eficiente es necesario resolverlo muchas veces pero ¿Quién necesita resolver el mismo problema muchas veces? La gran ventaja del Rank GA radica en que es siempre razonablemente eficiente sin la necesidad de hacer ajustes en él.

#### 4.5.2 Casos con menos neutralidad

Ahora se demuestra cómo es que, para los casos más difíciles, el Rank GA puede ser superior a cualquier configuración del SGA.

Usando una función que requiera más codones para ser codificada y en donde, en lugar de haber muchos términos cortos, haya pocos términos largos, se puede conseguir un paisaje de aptitud mucho más difícil.



**Figura 4-5** Curvas Características de Búsqueda del RGA y de varios casos del SGA con diferentes tasas de mutación (incluyendo la mejor posible) en el caso de la regresión de la función  $f(x) = 2\tanh(x+2)$ .

En la **Figura 4-5** se grafican los resultados obtenidos con la función  $f(x) = 2\tanh(x+2)$ . Un ejemplo de un fenotipo óptimo para esta función es  $+h+x+11h+1+x$ . El número de configuraciones óptimas es mucho menor que en el caso anterior por varias razones. Primero, por ser fenotipos más largos que requieren más codones activos o expresados en el fenotipo. Segundo, por tratarse solamente de 2 términos dando solamente dos posibilidades para conmutarlos. Tercero, dentro de cada término, es posible conmutar los unos y la  $x$  pero la  $h$  y

el signo + deben estar siempre en la misma posición haciendo a los genotipos óptimos más frágiles ante cambios genéticos. En consecuencia se tiene un paisaje de aptitud con mucho menor número de configuraciones óptimas distribuidas en picos mucho menos robustos que en la sección anterior.

Los resultados (**Figura 4-5**) revelan que, para el SGA, es siempre muy difícil hallar un óptimo incluso en el mejor de los casos. Por el contrario, el RGA es capaz de encontrarlo en menos de 1,000,000 de evaluaciones en aproximadamente el 60% de las repeticiones lo cual es sumamente eficiente si se considera que el espacio de búsqueda es de tamaño

$$2^{96} = 79,228,162,514,264,337,593,543,950,336 \approx 7.9 \times 10^{28}$$

La pendiente de la curva para el RGA permite aproximar cual sería el esfuerzo máximo para llegar al óptimo que debe ser algo quizás cercano a  $10^7$  evaluaciones.

## 4.6 Conclusiones

Se ha probado mediante un modelo con neutralidad que ésta juega un papel importante en la evolución genética al darle mayor robustez a las soluciones posibles y en general a todos los fenotipos. Las soluciones más robustas serán preferidas sobre las más frágiles lo que implica que, aún cuando se haya encontrado una solución óptima, la evolución continúa, dándose una competencia entre soluciones óptimas en la que ganan aquellas que sean más robustas ante los cambios provocados por los operadores genéticos. Además, en caso de que se tengan señales subóptimas, estas serán más importantes mientras más robustas sean haciéndose necesarias tasas de mutación mayores para poder explorar más allá de la vecindad de dichas señales.

Por otro lado, la similitud entre la estructura genética emergente en el modelo evolutivo usado en este capítulo con la estructura gramatical de los lenguajes hablados, permite analizar el sistema desde un punto de vista distinto y además se obtiene una explicación que abarca ambos sistemas. Se tienen diferentes grados de libertad efectivos a diferentes niveles. Las letras o sonidos básicos en el nivel más bajo se desarrollan en los individuos hasta que ya no son suficientes y se inicia el desarrollo de las palabras que son combinaciones de sonidos preferidas que amplían el número de posibles ideas que se pueden transmitir y finalmente se desarrollan secuencias de palabras preferidas sobre otras ya que siguen una estructura gramatical que permite también ampliar el número y la profundidad de los mensajes que pueden transmitirse entre los individuos.

La aplicación del Rank GA estudiado en el Capítulo 3 en este mapeo genotipo-fenotipo complejo demostró las buenas propiedades del algoritmo cuando el paisaje de aptitud presenta características tales como alta rugosidad, diferentes grados de neutralidad (en soluciones subóptimas y óptimas), decepción, fragilidad de los bloques constructores del óptimo, etc. También quedó claro que el SGA puede ser ajustado para ser eficiente en algunos problemas pero en otros simplemente no, y que el Rank GA es razonablemente eficiente tanto en problemas difíciles como fáciles.

## Notas

## Capítulo 5 **Mapeo Fenotipo-Aptitud complejo**

En los capítulos anteriores se estudiaron sistemas genéticos desde los más simples haciéndose cada vez más complejos al ir introduciendo más operadores (y sus variantes) además de irse agregando y combinando paisajes de aptitud básicos para llegar a paisajes complejos.

El siguiente paso es ahora estudiar mapeos fenotipo-aptitud complejos que sean más parecidos a los sistemas biológicos reales. ¿Porqué es necesario esto? Porque en la naturaleza no hay una forma explícita y directa de asignar valores de aptitud a los individuos sino que sus rasgos fenotípicos se traducen en una aptitud efectiva de manera indirecta y codependiente del estado de los demás individuos. El objetivo de este capítulo es entender mejor las interacciones presentes en el espacio de los rasgos fenotípicos en estos sistemas no triviales.

¿Que es entonces la aptitud en un sistema biológico? ¿Cómo se mide? En la naturaleza intervienen múltiples factores para determinar qué tan exitoso es un genotipo y no siempre es la especie físicamente más fuerte la que termina dominando a las demás. Entonces ¿qué rasgos fenotípicos hacen a una especie apta para sobrevivir en un medio ambiente dado? ¿bajo qué circunstancias es favorable desarrollar alguna capacidad específica? ¿existe alguna relación entre el desarrollo de unos rasgos fenotípicos y otros? Son algunas de las preguntas a las que se da respuesta en este capítulo en base a las observaciones hechas en un simulador de vida artificial que modela diferentes rasgos fenotípicos de los individuos.

Con este modelo se pretende mostrar cómo es que la naturaleza asigna un valor de aptitud a los individuos de manera indirecta y compleja y que al analizar los resultados es posible distinguir los grados de libertad efectivos, macroscópicos y microscópicos, que no son claros a simple vista. Se identifican los bloques constructores que contribuyen a una buena aptitud efectiva. Como se verá, estos bloques constructores surgen a partir de los rasgos fenotípicos como son la capacidad sensorial, de raciocinio, de metabolismo, etc. Uno de los objetivos es medir hasta qué grado están relacionados estos rasgos.

### 5.1 **El Sistema**

El sistema es uno en el que individuos de diferentes especies coexisten en un mismo espacio físico (geográfico) de dos dimensiones y compiten por el alimento disponible distribuido sobre dicho espacio. La energía vital de cada individuo se va gastando y mueren cuando se les agota por completo. Los individuos que logran metabolizar alimento incrementan su nivel de energía vital lo que les permite vivir más.



Uno de los rasgos fenotípicos que pueden tener los individuos es una estrategia fija para hacer uso de sus capacidades como son: metabólica, motriz, sensorial y de raciocinio, que están dadas también como rasgos fenotípicos del individuo. El uso de cada una de estas capacidades tiene un costo energético.

Entonces ¿Cuál será el conjunto de rasgos fenotípicos más adecuado para la supervivencia? Es claro que hay un compromiso entre el ahorro de energía y el uso adecuado de esta energía para que el individuo logre obtener alguna ventaja sobre los demás. ¿Cuándo conviene hacer uso de cual capacidad y en qué cantidad? La respuesta está en el medio ambiente en el que se encuentren los individuos. Es el medio el que ejerce presión para que se desarrollen unos u otros rasgos fenotípicos.

## 5.2 El modelo

Los individuos tienen cada uno un nivel de energía que, de ser consumida totalmente, provoca su muerte. Inicialmente todos los individuos tienen un mismo nivel de energía. Hay una pérdida de energía constante  $E_{\gamma t}$  por cada unidad de tiempo  $\Delta t$ . Los individuos pueden decidir una de tres acciones: *Comer*, *Observar* o *Moverse* en cada unidad de tiempo. Si deciden *Comer*, recuperan energía mediante el uso de su capacidad de metabolizar el alimento disponible en el lugar geográfico en el que están situados, sin rebasar un máximo predefinido de energía  $E_{max}$ . Si deciden *Observar* harán uso de su capacidad sensorial (con un costo energético proporcional a ésta) para obtener información de la cantidad de alimento a su alrededor. Si deciden *Moverse* harán entonces uso de su capacidad motriz (a un costo energético también proporcional a ésta) para cambiar su posición en alguna dirección que ellos decidan. La dirección de movimiento puede ser aleatoria o en función de la última observación hecha.

Se definen para cada individuo  $i$  las siguientes cantidades

$C_m^i$  Capacidad metabólica

$C_s^i$  Capacidad sensorial

$C_v^i$  Capacidad motriz

El uso de la capacidad  $r$  tiene un costo  $B_r^i$  por cada  $\Delta t$  en el nivel de energía del individuo  $i$  proporcional a la capacidad  $C_r^i$  en  $i$  dada por  $B_r^i = K_r C_r^i$  donde  $K_r$  es el costo por unidad de capacidad  $r$ .

La ecuación de energía para todo individuo  $i$  es

$$E_i(t + \Delta t) = E_i(t) - E_{\gamma t} + B_{[h_i(t)]}^i - A_{[h_i(t)]}^i$$

donde  $h_i(t)$  que es la capacidad que el individuo  $i$  ha decidido usar al tiempo  $t$ ,

$$A^i(t) = \begin{matrix} \min \{ C_m^i, f(x_i(t), y_i(t), t), E_{\max} - E_i(t) \} \\ 0 \\ 0 \end{matrix}$$

es un vector formado por la cantidad de energía que sería recuperada por el individuo  $i$  para cada una de las decisiones posibles: *Comer*, *Observar* o *Moverse*, en ese orden. Nótese que solamente la acción *Comer* tiene un valor mayor a cero.  $f(x_i(t), y_i(t), t)$  es la cantidad de alimento restante al tiempo  $t$  en la posición  $(x_i(t), y_i(t))$  que es la posición del individuo  $i$  al tiempo  $t$ .

Los individuos pueden tener capacidad de raciocinio o no. Cuando un individuo decide *Moverse* y tiene capacidad de raciocinio, lo hace hacia donde sabe que hay más alimento según la información dada por la última observación hecha. Si no tiene capacidad de raciocinio elige la dirección del movimiento de forma aleatoria.

Dos individuos no pueden estar en la misma posición al mismo tiempo y cuando alguno intenta ocupar la posición de otro, la posición será ganada por el individuo con mayor nivel de energía. Si el que solicita la posición pierde, simplemente se queda donde está. Si el que la solicita gana, desplaza al otro a la posición que el ganador tenía.

Cada individuo sigue una estrategia fija para elegir la siguiente acción a realizar en el tiempo  $t$  (comer, observar, moverse) que puede ser una de las siguientes

- ? Acción aleatoria
- ? Secuencia observar, moverse y comer
- ? Comer hasta agotar el alimento en la posición actual, luego moverse.
- ? Comer y observar hasta detectar más alimento en otra posición, luego moverse.
- ? Comer y observar hasta detectar igual o más alimento en otra posición, y moverse.

### 5.3 El Simulador

El simulador es un programa hecho en el paquete comercial Borland C++ Builder [ 9 ] que consta de un panel visible en pantalla en donde se desarrolla en tiempo real la vida artificial controlada para su análisis visual. Se genera además un archivo de resultados de cada corrida a partir del cual se hace un análisis estadístico y/o se crean gráficas para su análisis.

#### 5.3.1 Terreno

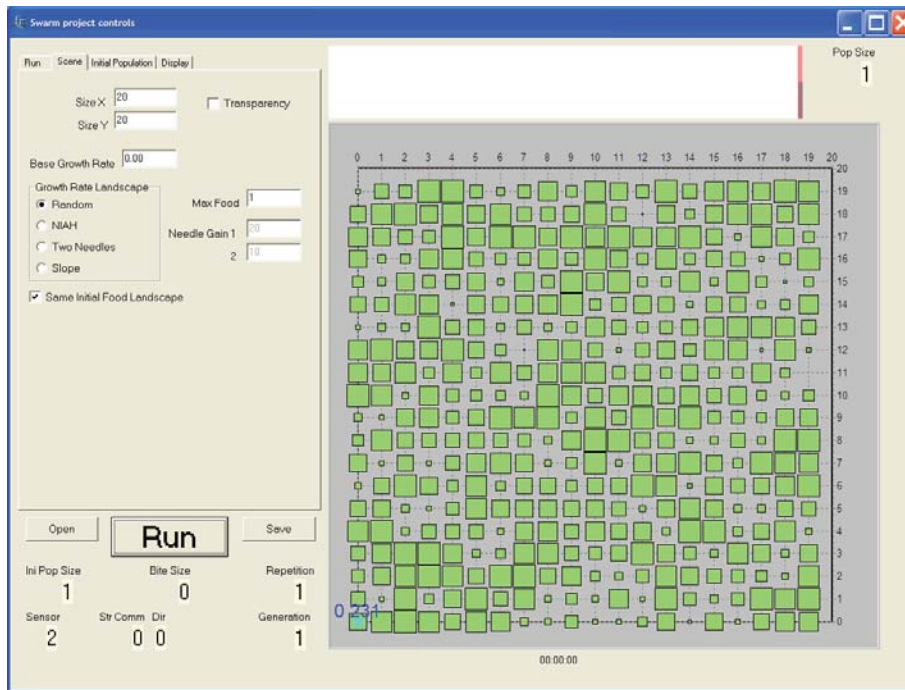
El terreno es un espacio rectangular plano de dos dimensiones formado por una malla de

casillas identificadas por sus coordenadas rectangulares. Los individuos no pueden salir de los límites de este terreno.

### 5.3.2 *Alimento*

En cada casilla del terreno hay una cantidad de alimento que puede ser consumida por los animales. La distribución inicial del alimento sobre el terreno puede hacerse de diferentes maneras: *Aleatoria*, *Aguja en un pajar*, *Dos agujas* o en *Pendiente* (ver **Figura 5-1**, **Figura 5-2** y **Figura 5-3**).

También puede especificarse una tasa de crecimiento del alimento para cada casilla. La distribución de estas tasas sobre el terreno es siempre la misma que la distribución inicial del alimento. El alimento crece solamente hasta alcanzar un nivel máximo posible.



**Figura 5-1** Distribución de alimento “Random”.

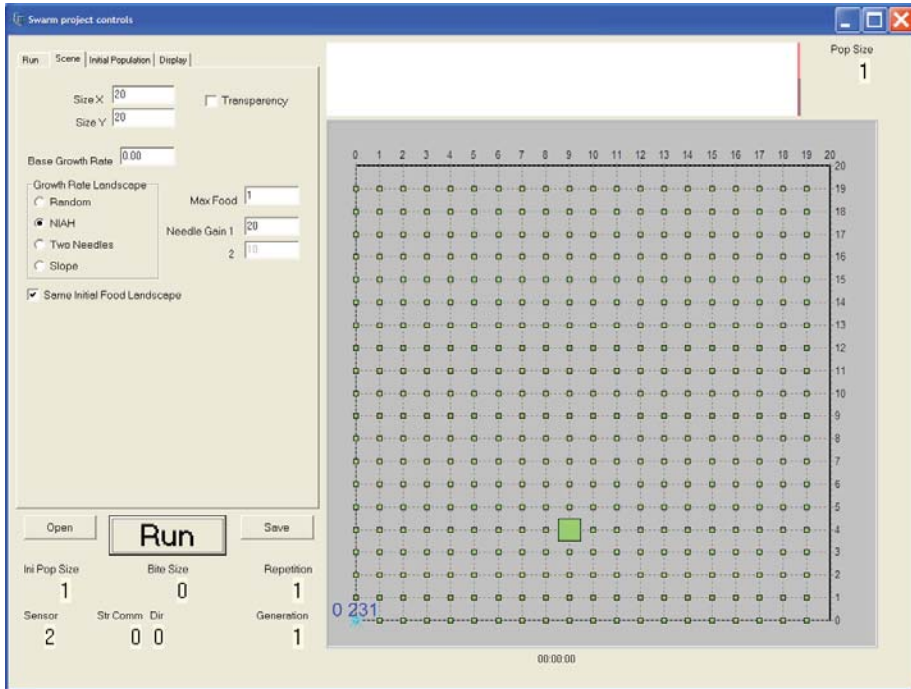


Figura 5-2 Distribución de alimento “NIAH”.

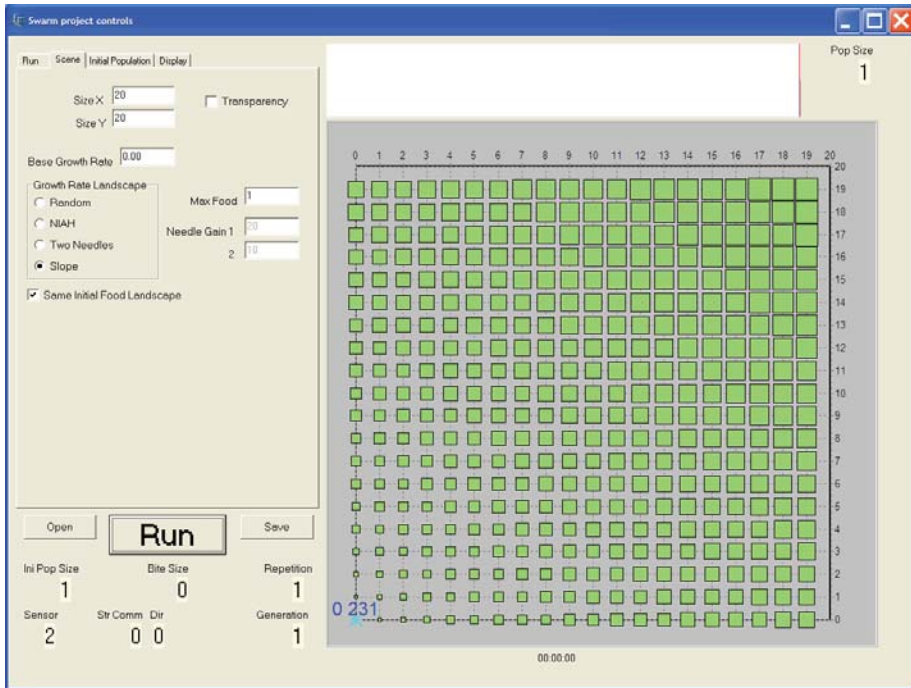


Figura 5-3 Distribución de alimento “Slope”.

### 5.3.3 *Especies*

Las posibles especies se distinguen por la combinación de una estrategia de comandos (CommStr) y una estrategia de dirección de movimiento (DirStr) o capacidad de raciocinio.

Las estrategias de comando pueden ser una de las siguientes:

- 0 Comando aleatorio
- 1 Comer/Observar/Mover
- 2 Comer y Observar hasta que ya no haya alimento en la casilla actual y entonces moverse.
- 3 Observar y si se detecta una casilla con más alimento entonces Mover sino Comer.
- 4 Observar y si se detecta una casilla con igual o más alimento entonces Mover sino Comer.

Hay dos capacidades de raciocinio posibles que son: el usar correctamente la información proporcionada por el sensor o ignorarla. Se definen entonces dos posibles estrategias de dirección de movimiento:

- 0 Dirección aleatoria
- 1 Hacia la posición en donde hay más alimento (o aleatoria si el sensor no da suficiente información)

Así se tienen 10 posibles especies combinando cada una de las estrategias de comandos con cada estrategia de dirección de movimiento.

Véase en el Apéndice 3 más detalles sobre las opciones posibles en el simulador.

## 5.4 Resultados

La métrica para hacer este análisis es el tiempo que tardan los individuos en morir sobre un terreno en el cual hay una cantidad fija de alimento que no se regenera. Se usan las distribuciones de alimento iniciales: *Random* (**Figura 5-1**), *NIAH* (**Figura 5-2**) y *Slope* (**Figura 5-3**).

En los experimentos que a continuación se analizan se fijaron algunos parámetros en el simulador (ver Apéndice 3) como son: InitialPopMode = Auto, MaxEnergy = 10, IniEnergy = 5, DeltaEnergy = 0.005 =  $E_{\gamma}$ , Duplication = false, SensorCost = 0.0003 =  $K_s$ , SizeX = 20, SizeY = 20, BaseGrowthRate = 0, MaxFood = 1, NeedleGain1 = 20, Transparency = false.

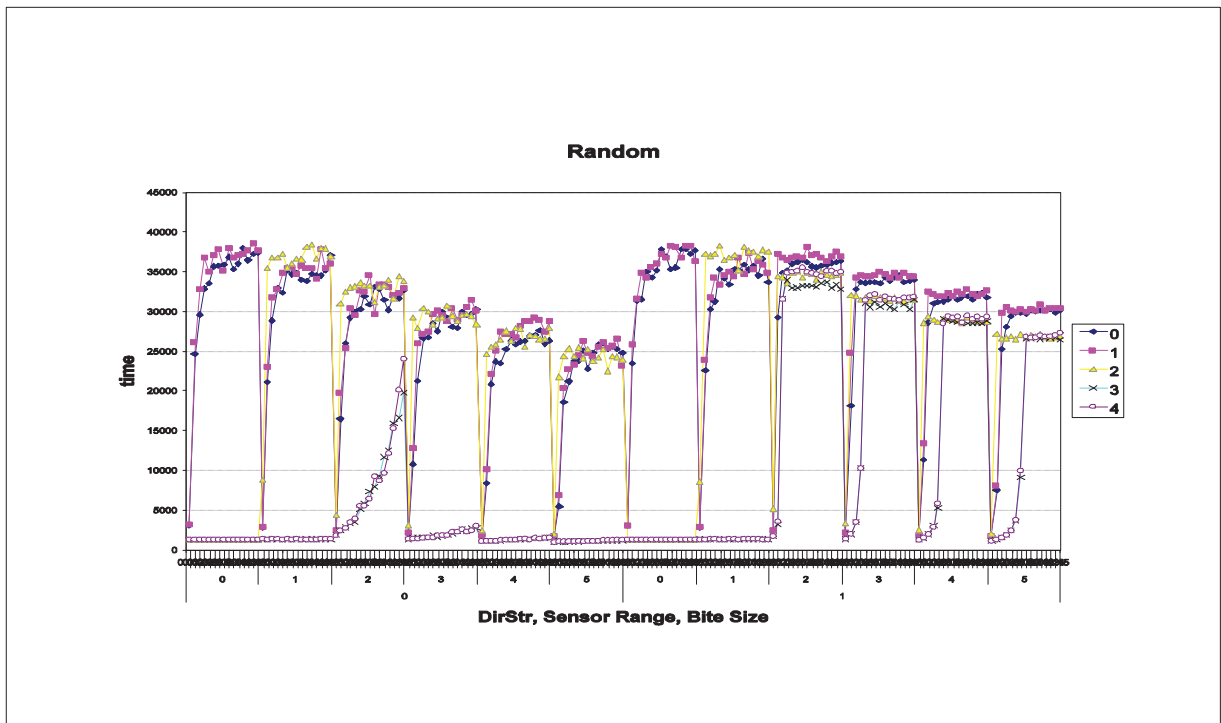
En los casos cuando se usa un rango de capacidad sensorial ( $C_s$ ), éste fue de 0 a 5, y en el caso de capacidad metabólica (o tamaño de la mordida), se uso desde 0.01 a 0.15 ( $C_m$ ). Con  $K_m = 1$ . La capacidad motriz se fijó en 1 y el costo de la capacidad motriz se fijó en  $K_m = 0$ .

### 5.4.1 Individuos solos

Para tener un punto de comparación con los experimentos más complejos se analizan los casos en donde hay un solo individuo en cada una de las distribuciones de alimento *Random*, *NIAH* y *Slope*.

Se midió el tiempo que el individuo tardó en morir y se hicieron repeticiones variando *BiteSize*, *SensorRange*, *CommandStrategy* y *DirectionStrategy*.

En la **Figura 5-4** se grafica, para la distribución de alimento *Random*, el tiempo que logra vivir el individuo como función de su capacidad metabólica (de 0.01 a 0.15), para cada capacidad sensorial (de 0 a 5), cada estrategia de dirección (0 izquierda y 1 derecha) y cada estrategia de comandos (0 a 4 en las leyendas).



**Figura 5-4** Tiempo que tarda en extinguirse un individuo solo usando cada estrategia en función de su capacidad de raciocinio (*DirStr*), sensorial (*Sensor Range*) y metabólica (*Bite Size*) en la distribución de alimento "Random".

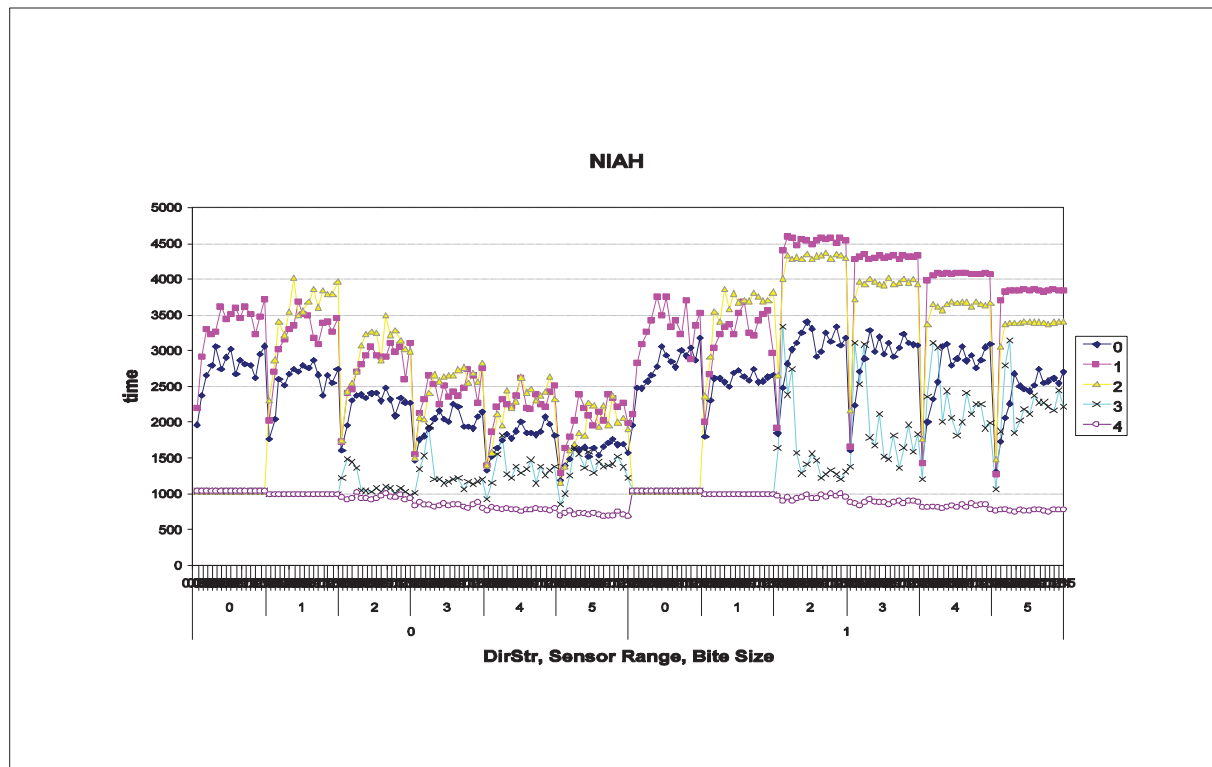
Se observa que no hace falta desarrollar sensor para sobrevivir más tiempo, ya que el alimento está distribuido irregularmente sobre todo el terreno y, por lo tanto, no es difícil hallar buenas fuentes de alimento haciendo caminatas aleatorias, es más, el sensor solamente causa costos de energía haciendo más corta la vida del individuo. Lo que sí ayuda ligeramente es usar bien la información ( $DirStr = 1$  a la derecha) cuando hay un sensor capaz de ver más allá de la casilla propia ( $C_s > 1$ ), es decir, el moverse hacia donde hay más alimento en lugar de tomar una dirección aleatoria.



Se observa también que hay que tener un mínimo de capacidad metabólica (*Bite Size*) para compensar la pérdida de energía en los períodos entre una acción de *Comer* y otra que es cuando se hacen observaciones y movimientos. Para valores altos de capacidad metabólica ya no es muy útil seguirla aumentando ya que pronto se llega a un punto en el que el alimento restante en las casillas es menor a esta capacidad haciéndose imposible metabolizar más alimento que el que está presente.

La estrategia 2 requiere sensor 1 para ser óptima ya que solo necesita la información de la casilla propia (y hacer uso de más capacidad sensorial no ayuda como se explicó arriba) y las estrategias 3 y 4 requieren al menos sensor 2 y por lo tanto, estas últimas, requieren una mayor capacidad metabólica para recuperar más rápidamente la energía usada por el sensor. Las estrategias 3 y 4 son más selectivas en cuanto a los puntos en los que deciden comer. Requieren que estos puntos sean máximos locales. Por esto, cuando  $DirStr = 0$ , requieren mucho más capacidad metabólica ya que, cuando deciden moverse, lo hacen hacia cualquier parte y, por lo tanto, puede ser que se acerquen o alejen del punto donde decidirían comer lo que se traduce en caminatas aleatorias largas sin comer.

Los grados de libertad efectivos para obtener un individuo apto, en este caso, son la autosuficiencia alimentaria mediante una capacidad metabólica suficiente, el consumo de energía mínimo posible mediante una capacidad sensorial baja y el tener una estrategia que con suficiente frecuencia ordene alimentarse y que no tenga largos períodos sin hacerlo.

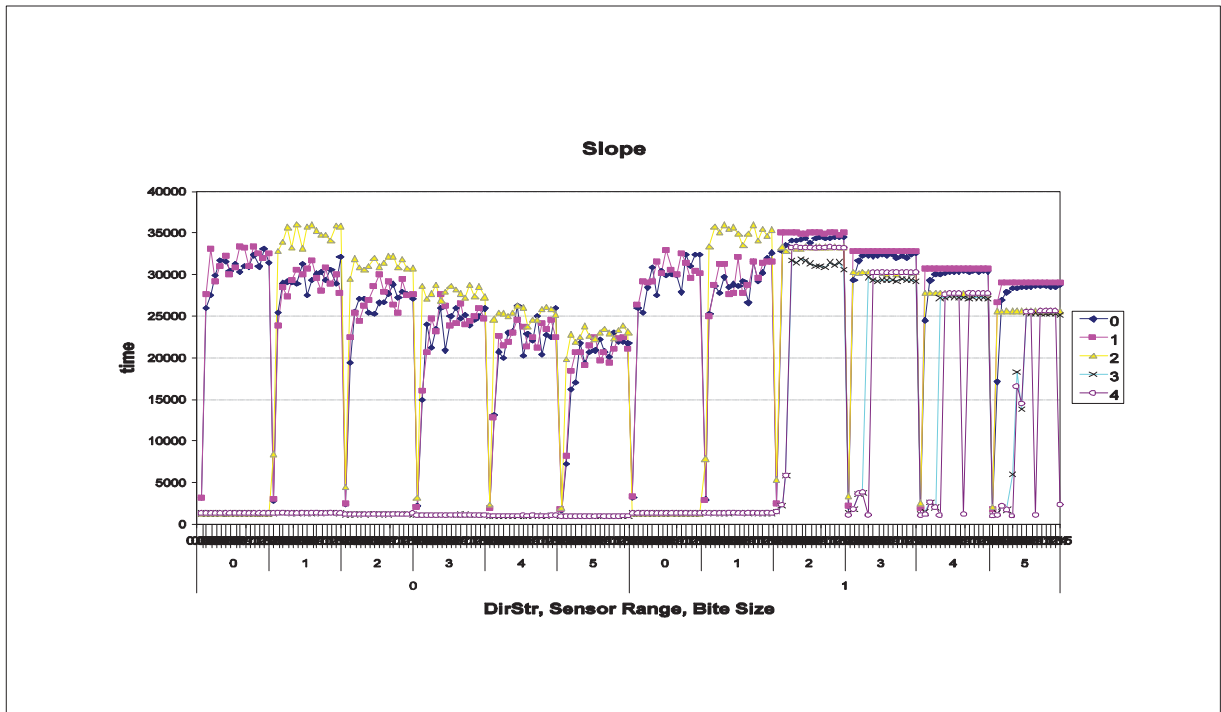


**Figura 5-5** Tiempo que tarda en extinguirse un individuo solo usando cada estrategia (en leyendas) en función de su capacidad de raciocinio (*DirStr*), sensorial (*Sensor Range*) y metabólica (*Bite Size*) en el paisaje “NIAH”.

En el caso de distribución de aguja en un pajar (*NIAH*) (**Figura 5-5**) es notable que la estrategia 4 es muy mala ya que siempre encuentra una casilla con igual cantidad de alimento y decide moverse todo el tiempo sin comer. La estrategia 3 tiene problemas también cuando el alimento se agota en alguna región ya que decide seguir comiendo si no detecta casillas con más alimento lo cual es imposible si ya no hay alimento en su rango sensorial, y por esto se observa una mejoría al incrementar la capacidad sensorial. En este caso y con capacidad sensorial baja, un tamaño de mordida pequeño es el óptimo ya que así no se agota el alimento de las casillas de una sola mordida (en este paisaje la aguja tiene 1 de alimento y la paja tiene  $1/20 = 0.05$ ) logrando sobrevivir mucho más tiempo. Al aumentar la capacidad sensorial aumenta la probabilidad de hallar la aguja y por lo tanto aumenta también el tiempo de supervivencia.

Para las demás estrategias, se nota que el mejor sensor es 2 cuando se usa la dirección 1. Con esta configuración es posible hallar la zona del paisaje con más alimento (la aguja) al más bajo costo. La estrategia 0 es peor que las estrategias 1 y 2 ya que en ocasiones toma malas decisiones como hacer movimientos sin haber recabado información previa y por lo tanto aleatorios o hacer movimientos cuando está en la aguja en lugar de comer.

El grado de libertad agregado es la capacidad para detectar la zona con más alimento y de moverse hacia ella al menor costo y se acentúa la capacidad de raciocinio como grado de libertad importante.



**Figura 5-6** Tiempo que tarda en extinguirse un individuo solo usando cada estrategia en función de su capacidad de raciocinio (*DirStr*), sensorial (*Sensor Range*) y metabólica (*Bite Size*) en el paisaje “Slope”.



Cuando la distribución inicial de alimento es *Slope* (**Figura 5-6**) la situación es parecida a *NIAH* en cuanto a que se debe hallar la zona del paisaje que tiene más alimento pero ahora hay señales en todo el paisaje que llevan a esta zona. En los resultados hay diferencias sutiles como por ejemplo que la estrategia de comando 0 es prácticamente igual a la estrategia 1 ya que, eventualmente, las dos llegan a la zona de más alimento y terminan aprovechándolo todo. Las estrategias 3 y 4 ya no son tan malas cuando la dirección es hacia lo mejor que se conoce ( $DirStr = 1$ ) por la misma razón de que pueden llegar a la mejor zona del terreno ya que no se estancan en zonas planas, aunque en casos aislados, se generan algunas zonas planas que sí les afectan y este es el caso para ciertos tamaños de mordida que son múltiplos de 0.05. La razón es que esta cantidad es precisamente la diferencia que hay entre las cantidades iniciales de alimento de una casilla a otra (para formar la pendiente).

En dirección 0 (aleatoria) la estrategia 3 es peor que en *NIAH* ya que en este paisaje solamente hay una casilla en donde puede comer y el resto del paisaje le dice que debe moverse (contrario al caso *NIAH*) pero como lo hace aleatoriamente le cuesta mucho encontrar dicha casilla y en la mayoría de los casos muere antes de hacerlo.

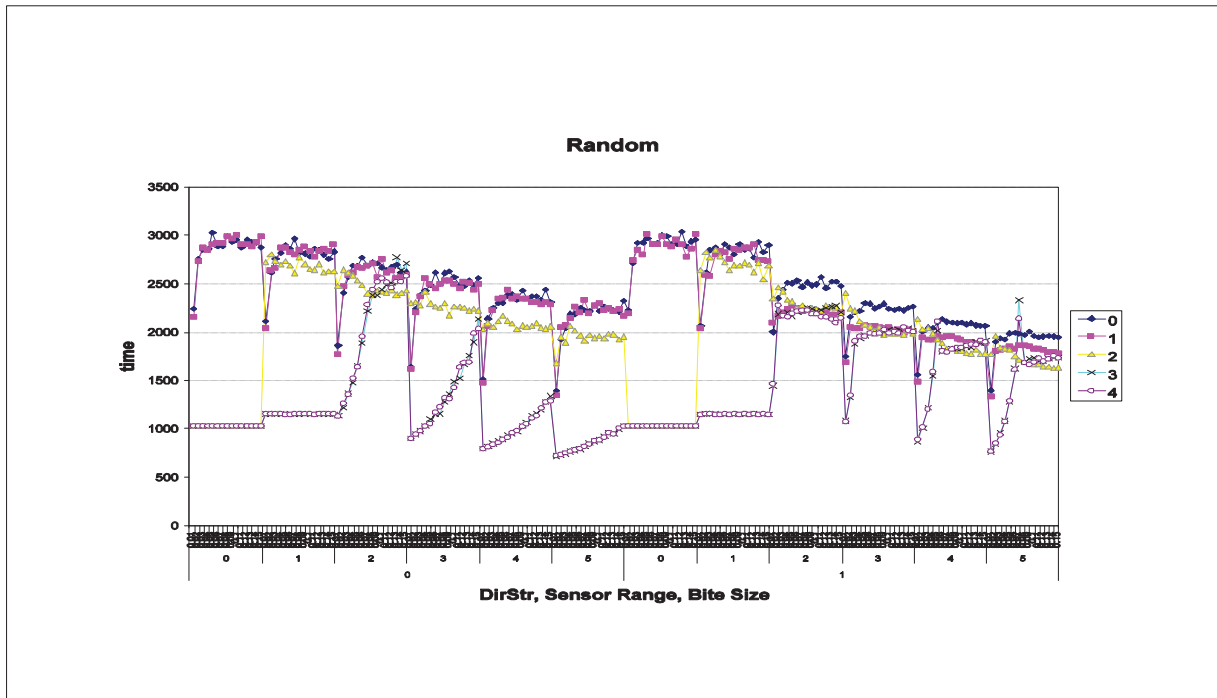
#### 5.4.2 ***Especies puras. 30 individuos iguales***

Ahora los experimentos consistieron en hacer 10 corridas con poblaciones de 30 individuos idénticos para cada especie y para cada distribución del alimento. Se midió el tiempo promedio que tarda cada especie en extinguirse por completo. Aquí se verá cuales características favorecen a la especie como grupo y no necesariamente a los individuos.

En el caso *Random* (**Figura 5-7**) se observa que, de igual manera que cuando se tenía un solo individuo, el desarrollo del sensor no ayuda en general. Sin embargo para  $DirStr = 1$  y sensor 2 se nota una perdida grande de capacidad de supervivencia del grupo de individuos. Nótese que para un individuo solo hay, en este caso, una ligera mejora. Esta mejora individual se traduce en una desventaja como grupo ya que todos tienen esa ventaja permitiéndoles vivir más tiempo y por lo tanto consumir más alimento lo que finalmente reduce la probabilidad de que al menos uno de ellos logre vivir mucho tiempo.

Un efecto similar se observa claramente en el caso  $DirStr = 1$ , sensor 5 y las estrategias 3 y 4. Aquí el mejor tamaño de mordida es uno suficientemente grande como para que los individuos tengan probabilidad de sobrevivir, pero no tan grande para que solo pocos sobrevivan. Así, la probabilidad de que haya alta mortandad al inicio es alta pero también hay alta probabilidad de que unos pocos individuos sobrevivan quedándose en un terreno con mucho alimento disponible todavía.

En el caso *NIAH* (**Figura 5-8**) se pierde el efecto visto cuando se tenía un individuo ya que ahora con muchos individuos es altamente probable que alguno encuentre la aguja. Si se ponen menos individuos o si se hace el terreno más grande podría verse una preferencia por un sensor mayor a cero por ejemplo, para así tener mayor probabilidad de encontrar esta aguja en el pajar. Por lo demás no hay cambios cualitativos importantes.



**Figura 5-7** Tiempo que tardan en extinguirse 30 individuos idénticos usando cada estrategia en función de su capacidad de raciocinio (DirStr), sensorial (Sensor Range) y metabólica (Bite Size) en el paisaje “Random”.

El grado de libertad efectivo es aquí la eliminación de los competidores para concentrar los recursos en pocos individuos.

Para el caso de distribución *Slope* (Figura 5-9) se da un efecto interesante. Además de los efectos vistos en *NIAH*, se tiene que las estrategias 3 y 4 tienen mejor desempeño y en algunos casos son las mejores. Esto se debe también a que es un solo individuo el que sobrevive quedándole mucho alimento disponible. Cuando DirStr = 0 los individuos no se dirigen hacia donde podrían comer y mueren antes de poder hacerlo, solo uno o dos lo logran y se come(n) todo el alimento solo(s). Cuando DirStr = 1 y la estrategia es 3, sí se dirigen hacia el alimento pero todos van juntos hacia la misma zona generando zonas donde se agota el alimento. Si un individuo entra en una zona donde no hay alimento se atora y muere. Esto deja a muchos atorados y solo unos pocos tienen la suerte de seguir sobre zonas con mucho alimento.

Lo anterior implica que no siempre los individuos más aptos físicamente son los que tienen más aptitud efectiva. La aptitud es una característica indirecta en la que se toma en cuenta no solamente que tan buenos son los individuos para hacerse de alimento sino que tan buenos son como especie para garantizar su supervivencia. En este caso se identifica que el grado de libertad efectivo es la eliminación de los competidores, es decir, que la especie que elimina a muchos de sus miembros en beneficio de unos pocos resulta ser la mejor ya que los recursos disponibles se concentran en esos pocos sobrevivientes permitiéndoles vivir más tiempo. Para lograr esto, se requiere que los individuos no sean muy capaces de hallar y consumir el alimento, sino que, al ser muchos, tengan dificultad para sobrevivir y que al ser pocos, no tanta.

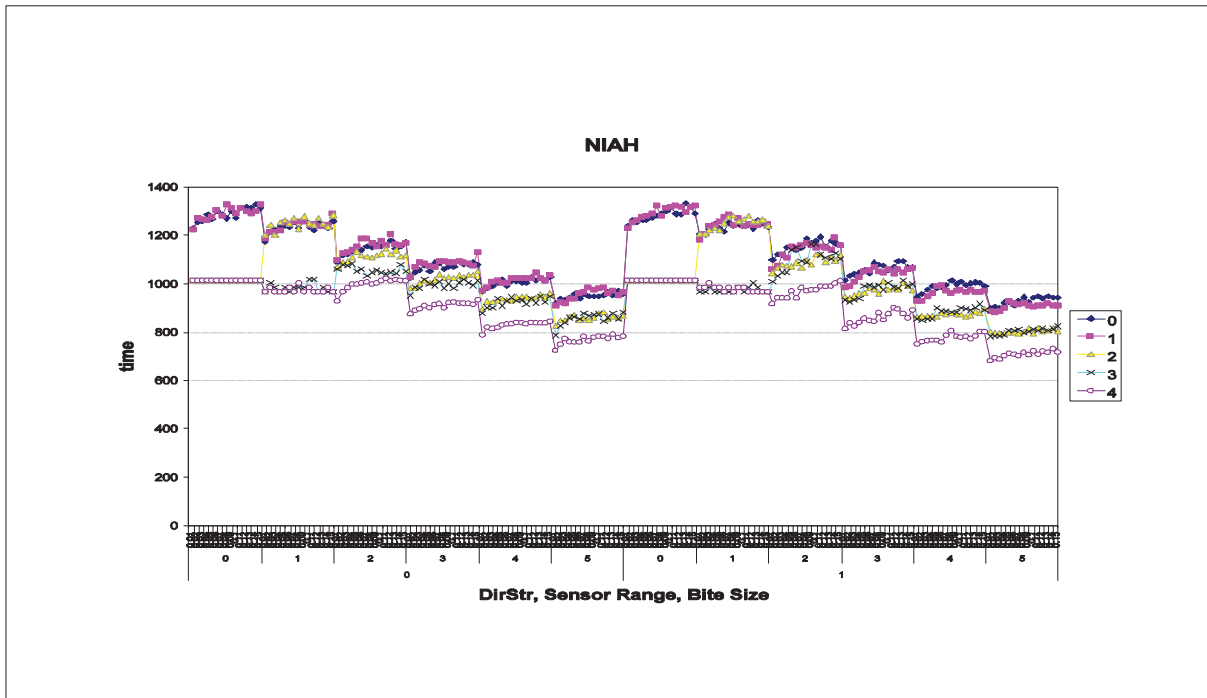


Figura 5-8 Tiempo que tardan en extinguirse 30 individuos idénticos usando cada estrategia en función de su capacidad de raciocinio (DirStr), sensorial (Sensor Range) y metabólica (Bite Size) en el paisaje “NIAH”.

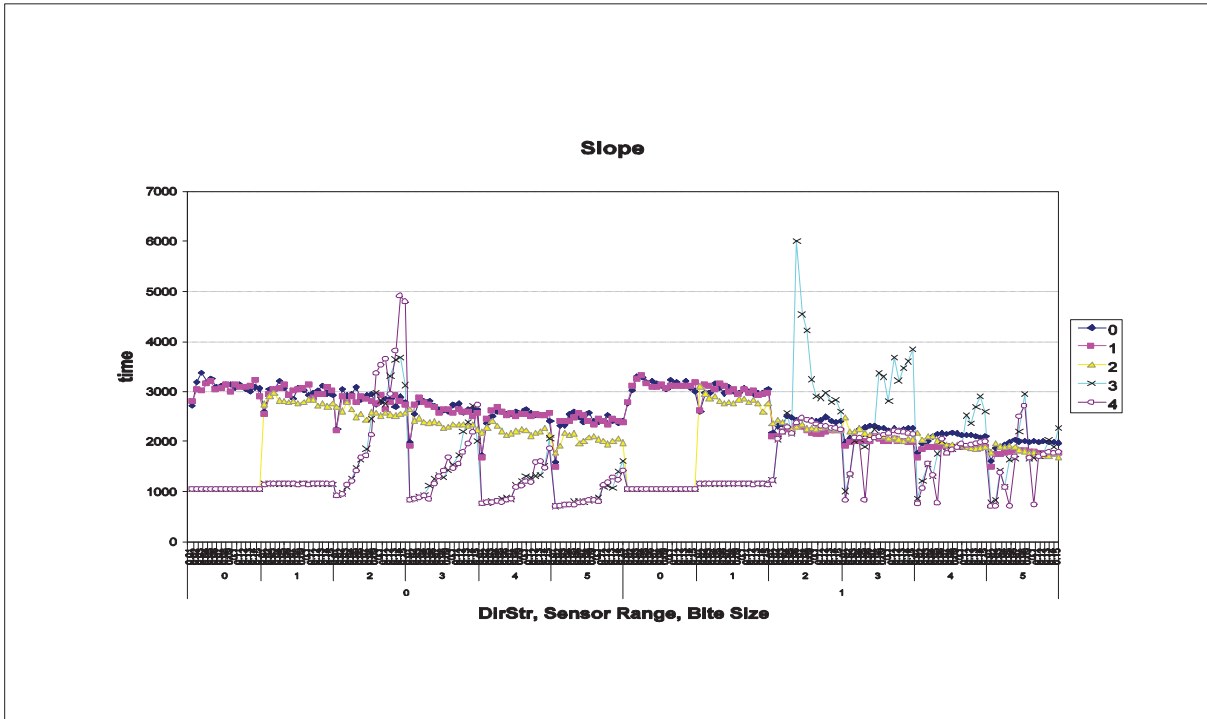


Figura 5-9 Tiempo que tardan en extinguirse 30 individuos idénticos usando cada estrategia en función de su capacidad de raciocinio (DirStr), sensorial (Sensor Range) y metabólica (Bite Size) en el paisaje “Slope”.

### 5.4.3 *Misma estrategia con diferentes capacidades*

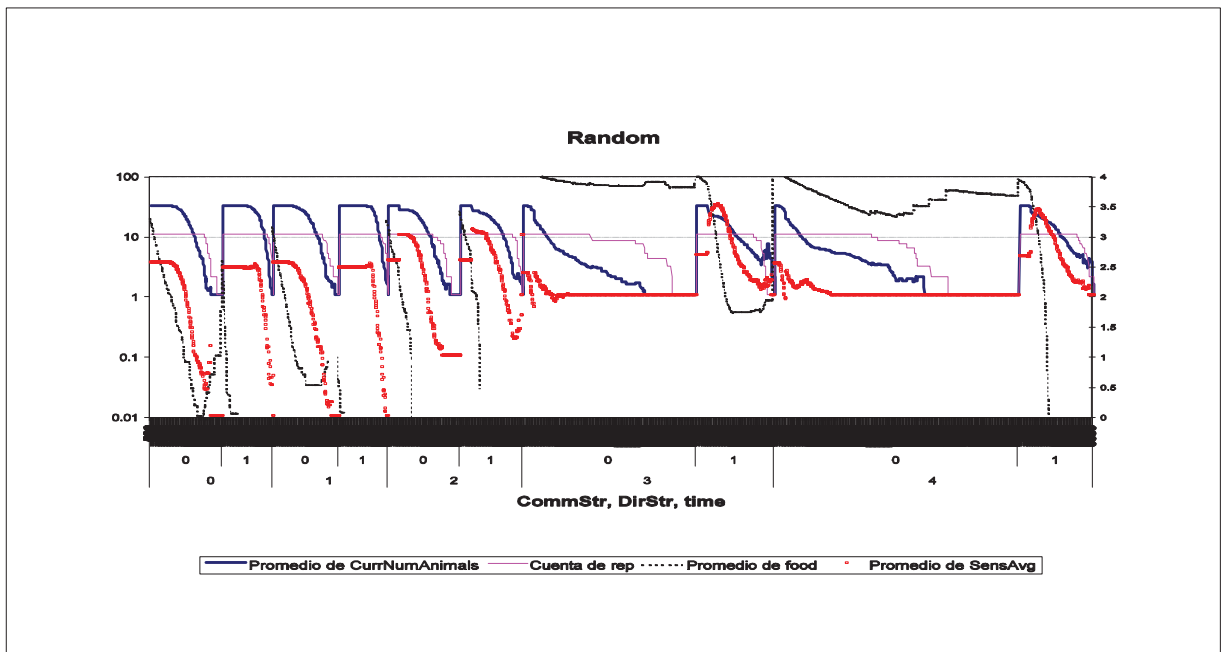
En la sección anterior se usaron poblaciones con individuos idénticos en todas sus capacidades y estrategias. Ahora surge la pregunta de ¿Qué tan bueno es distinguirse del grupo desarrollando capacidades? ¿Son siempre los individuos con más capacidad los más aptos? Para ver esto se correrán poblaciones con individuos cuyas capacidades son distintas.

#### 5.4.3.1 *Diferente capacidad sensorial, misma capacidad metabólica*

Para analizar estos casos se muestran en las gráficas, como función del tiempo y para 10 repeticiones:

- ? Número promedio de individuos vivos.
- ? Número de repeticiones con al menos un individuo vivo.
- ? Promedio de alimento restante.
- ? Promedio de capacidad sensorial en la población.

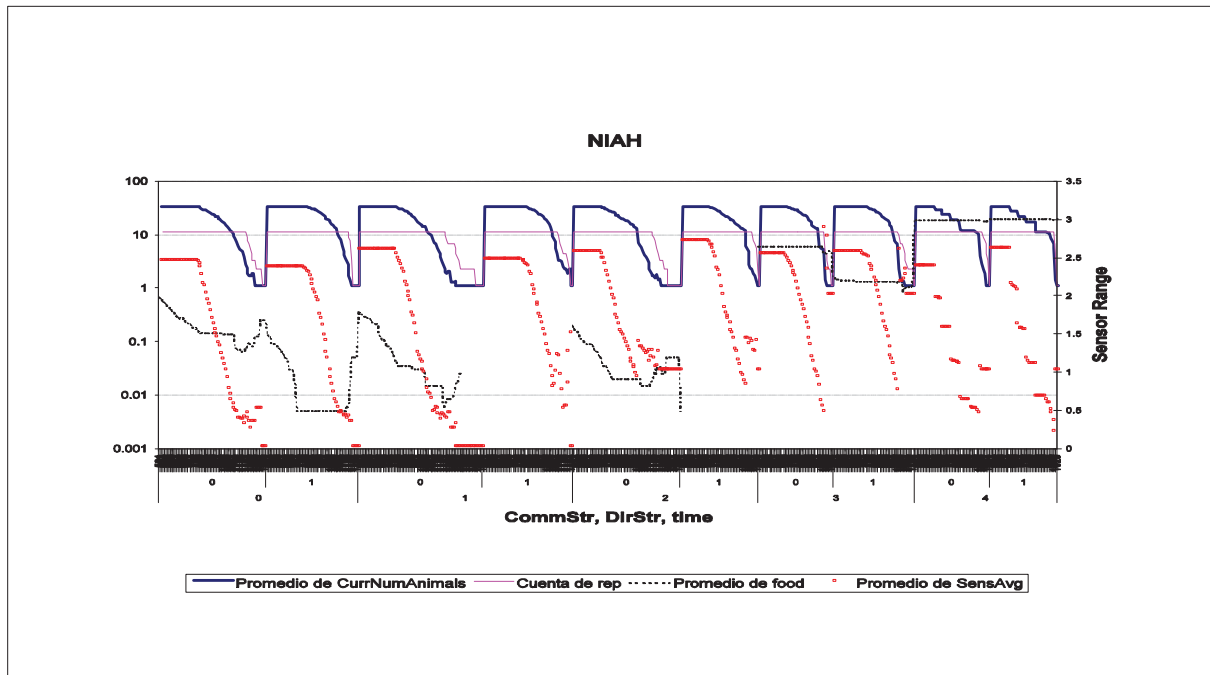
Se fijó Bite Size en 0.15 y se asignó aleatoriamente a los 30 individuos capacidades sensoriales entre 0 y 5.



**Figura 5-10** Resultados para cada estrategia de comando/dirección con poblaciones iniciales de 30 individuos de diferentes capacidades sensoriales en el paisaje “Random”.

En el paisaje Random (**Figura 5-10**) se ve que, en las estrategias 0 y 1, los individuos que dominan son los que tienen menor capacidad sensorial lo que es razonable ya que un comportamiento equilibrado entre comer y moverse es suficiente para hallar zonas con alimento sin gastar energía en un sensor. En la estrategia 2 el sensor preferido es 1 ya que esta

estrategia requiere saber cuando se ha agotado el alimento en su casilla actual y luego moverse aleatoriamente a otra zona. Para las estrategias 3 y 4 el sensor preferido es 2 para poder decidir si van a moverse o a comer sin quedar atorados.



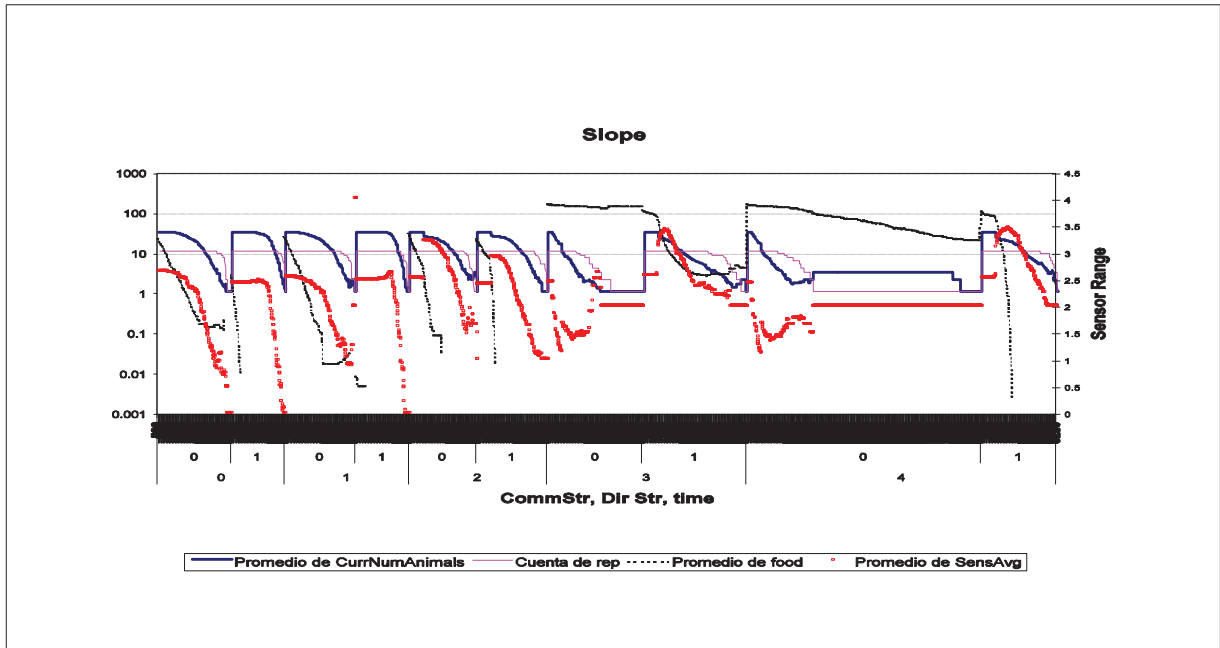
**Figura 5-11** Resultados para cada estrategia de comando/dirección con poblaciones iniciales de 30 individuos de diferentes capacidades sensoriales en el paisaje “NIAH”.

Para el paisaje NIAH (**Figura 5-11**) se ve que en ningún caso los individuos sobreviven mucho tiempo y es debido a que en el paisaje no hay señales que los lleven hacia el alimento, entonces, es cualquier individuo que esté inicialmente cerca de la aguja el que la encuentra sin importar cual sea su capacidad sensorial. Después de consumir el alimento en la aguja ya no queda mucho alimento en el resto del terreno porque ya fue consumido por los demás animales y por lo tanto el que encontró la aguja ya no puede sobrevivir mucho más. La ventaja adquirida por el que halló la aguja depende únicamente del costo energético de su capacidad sensorial y es por eso que la capacidad sensorial preferida es cero ya que esta le da la mayor ventaja. Los primeros en morir son siempre los que tienen más capacidad sensorial.

En el paisaje Slope (**Figura 5-12**) se tienen más o menos los mismos resultados para las estrategias 0 y 1 que en NIAH pero se desarrolla un sensor ligeramente mayor cuando la dirección es no aleatoria porque ahora sí rinde fruto la información que da el sensor. Para la estrategia 2 se ve que es preferible un sensor mayor a 1 cuando DirStr = 0, e igual a 1 cuando DirStr = 1. Esto se debe a que, en DirStr = 1, los individuos con sensor grande tienden a quedar varados cuando detectan más alimento en casillas que ya están siendo ocupadas por otros que están comiendo en ese lugar perdiendo energía mientras siguen tratando de llegar hasta allí y consecuentemente debilitándose para cuando el alimento escasea. El promedio de la capacidad sensorial disminuye ya que son éstos los primeros que van muriendo. En DirStr = 0 no pasa esto porque los individuos no van directamente a donde hallaron más alimento y por

consiguiente pueden sobrevivir mejor aquellos con sensor un poco mayor.

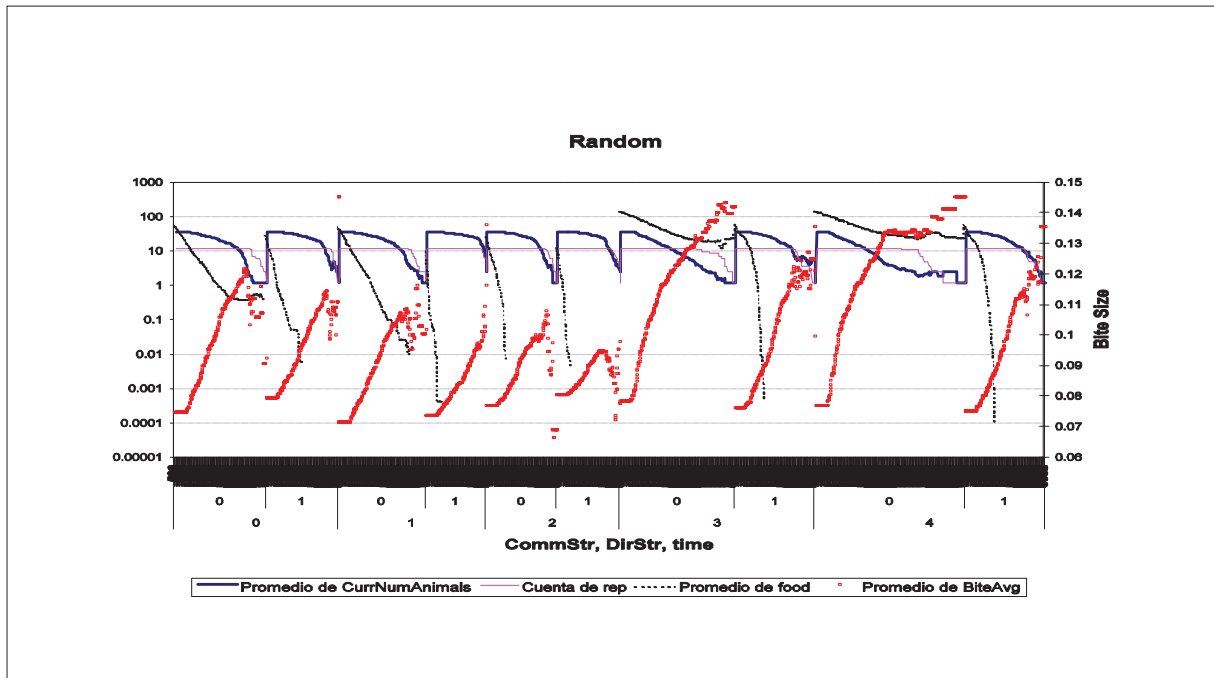
Se observa que en las estrategias 3 y 4  $\text{DirStr} = 0$  se prefieren sensores de capacidad cercana a 1 y con  $\text{DirStr} = 1$  capacidades un poco por arriba de 2. Con  $\text{DirStr} = 0$ , los individuos con sensores mayores tienden a ser muy inestables y comen menos llegando incluso a vivir menos que los de sensor 1 que no se mueven. Con  $\text{DirStr} = 1$  tienden a estorbarse y, cuando escasea el alimento, se vuelven malos por el costo del sensor y mueren antes, por lo tanto los de sensor 2 mueren al último. Los de sensor 0 no comen ni se mueven y los de sensor 1 no se mueven y comen solo inicialmente así que se mantienen quietos esperando a morir.



**Figura 5-12** Resultados para cada estrategia de comando/dirección con poblaciones iniciales de 30 individuos de diferentes capacidades sensoriales en el paisaje “NIAH”.

#### 5.4.3.2 Diferente capacidad metabólica, misma capacidad sensorial

Aquí se ponen ahora a competir individuos con diferentes capacidades metabólicas. Para el caso de paisaje Random (**Figura 5-13**), en las estrategias 3-0 y 4-0 es en donde ganan los individuos con mordida más grande. Esto es porque la mordida grande deja más dividendos en esas estrategias ya que logran acumular energía lo que les permita hacer caminatas aleatorias más largas sin comer. Las caminatas aleatorias más largas les dan mayor probabilidad de encontrar sitios para comer. El caso en donde la mordida que gana es menos grande es la estrategia 2 ya que con esta estrategia siempre se come todo el alimento disponible antes de moverse y, por lo tanto, solo es necesario metabolizar la cantidad necesaria para saber si ya se acabó el alimento en esa posición (usar el sensor) y acumular un poco más para eventualmente moverse a una casilla cercana.



**Figura 5-13** Resultados para cada estrategia de comando/dirección con poblaciones iniciales de 30 individuos de diferentes capacidades metabólicas en el paisaje “Random”.

En el paisaje NIAH (**Figura 5-14**) no hay sesgo por causa del tamaño de la mordida. El promedio de tamaño de mordida tiende siempre a valores no muy grandes. Esto se debe a que hay poco alimento en las casillas (excepto en una) y una mordida grande no puede aprovecharse ya que no puede morder más de lo que hay. En todos los casos hay un individuo que tiene la suerte de iniciar cerca de la aguja y la encuentra, independientemente de su capacidad metabólica. Los demás se quedan sin ventaja posible ya que solo hay una aguja. Esto reparte, sobre las repeticiones, la ventaja de la aguja entre todas las estrategias quedando nulificada después de muchas repeticiones.

Para el caso Slope (**Figura 5-15**) como en el caso Random se tiene una presión selectiva por tamaños de mordida grandes especialmente en las estrategias 3 y 4. Esto es porque, aunque para estas estrategias solo hay un punto inicialmente en el que comerían, eventualmente algún individuo (cuya mordida puede ser cualquiera así que no hay sesgo por algún tamaño) llega y al comer se inicia una cadena de cambios en el paisaje que lo convierten prácticamente en Random dándose los mismos efectos que en ese caso en cuanto al tamaño de la mordida.

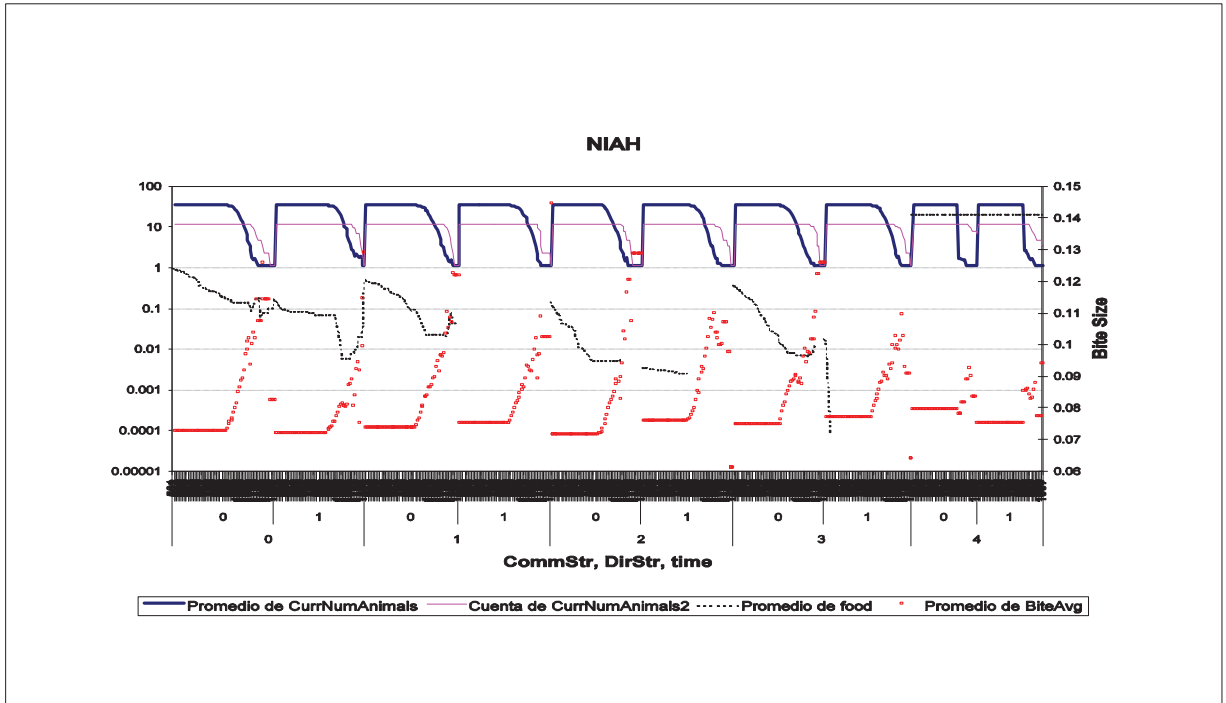


Figura 5-14 Resultados para cada estrategia de comando/dirección con poblaciones iniciales de 30 individuos de diferentes capacidades metabólicas en el paisaje “NIAH”.

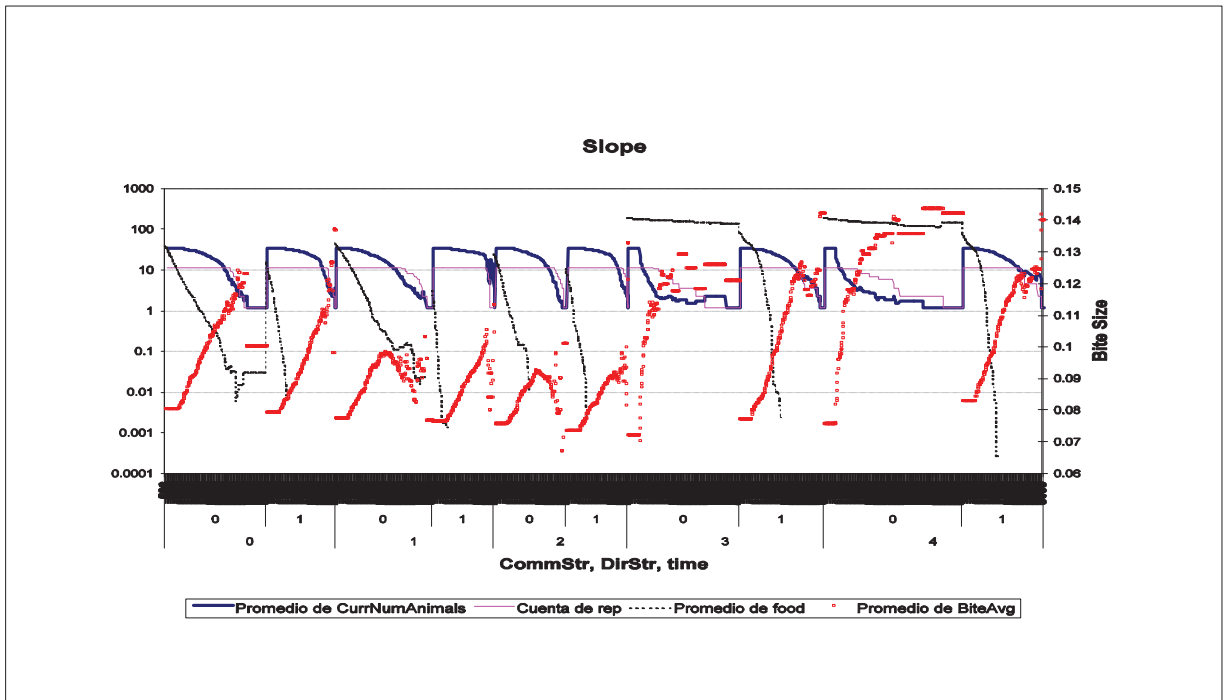


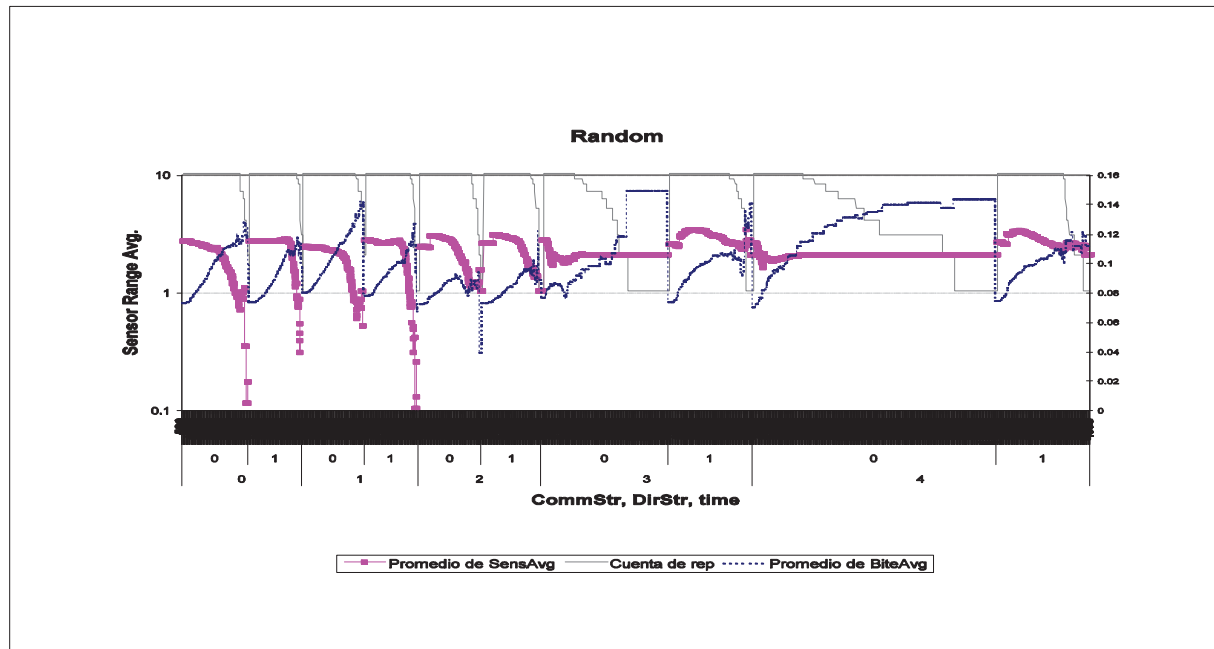
Figura 5-15 Resultados para cada estrategia de comando/dirección con poblaciones iniciales de 30 individuos de diferentes capacidades metabólicas en el paisaje “Slope”.



5.4.3.3 *Diferente capacidad sensorial y metabólica en cada individuo*

Aquí se ve como es que influyen entre sí el tamaño de la mordida y la capacidad sensorial al poner 30 individuos con diferentes valores de cada una de estas capacidades.

En general se ven los mismos efectos en la capacidad sensorial que cuando se tenía una mordida idéntica para todos. La diferencia está en el tamaño preferido de mordida que es ligeramente menor en las estrategias 2-0, 2-1, 3-1 y 4-1 debido a que es más fuerte la selección por aquellos con una capacidad sensorial adecuada que por aquellos con una muy buena capacidad metabólica descartándose así a algunos individuos con buena capacidad metabólica pero mala capacidad sensorial (mala en el sentido efectivo no absoluto). Véase la **Figura 5-16** la **Figura 5-17** y la **Figura 5-18** en donde se grafica el promedio de la capacidad sensorial, el número de repeticiones con al menos un individuo vivo y el promedio del tamaño de la mordida.



**Figura 5-16** Resultados para cada estrategia de comando/dirección con poblaciones iniciales de 30 individuos de diferentes capacidades sensoriales y metabólicas en el paisaje “Random”.

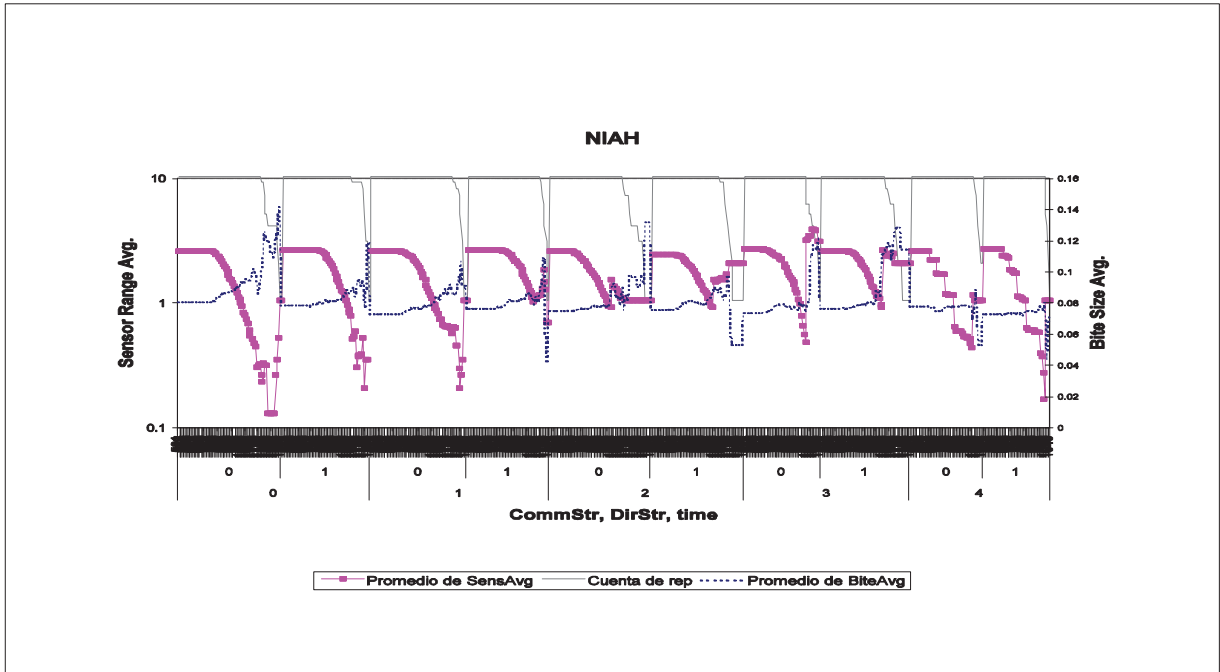


Figura 5-17 Resultados para cada estrategia de comando/dirección con poblaciones iniciales de 30 individuos de diferentes capacidades sensoriales y metabólicas en el paisaje “NIAH”.

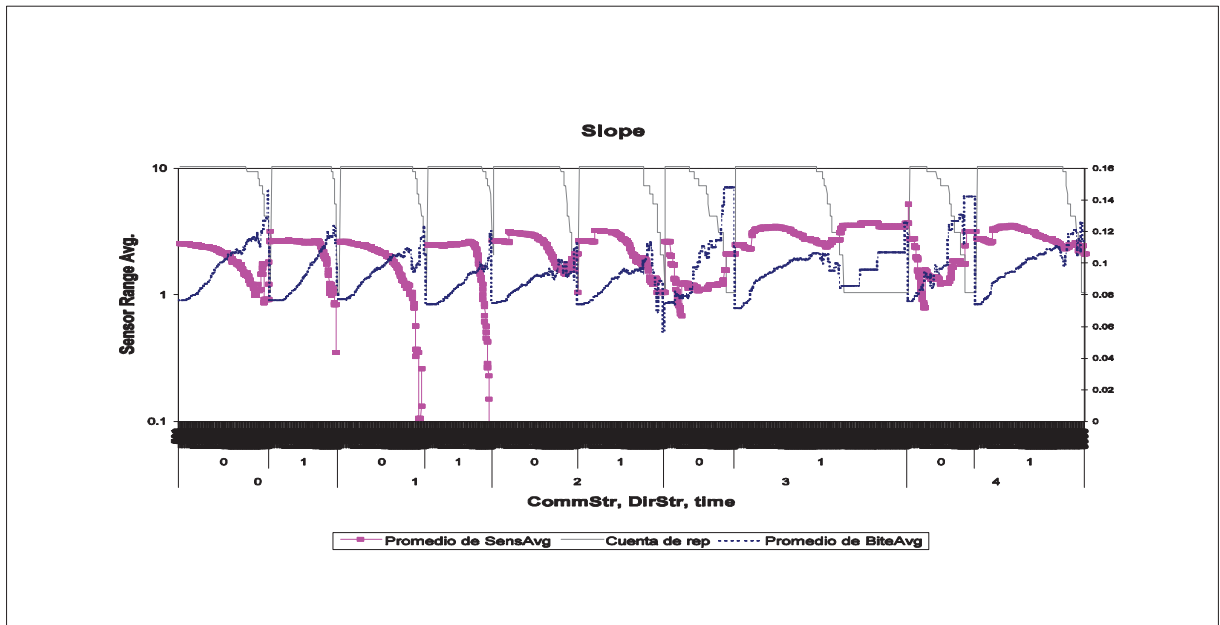


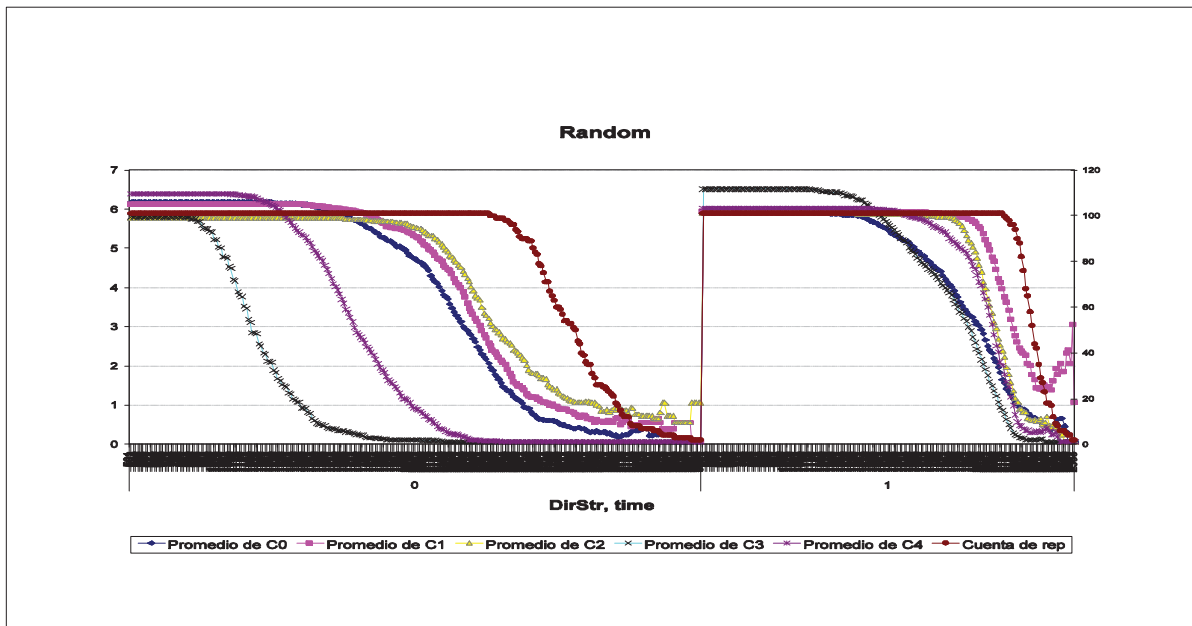
Figura 5-18 Resultados para cada estrategia de comando/dirección con poblaciones iniciales de 30 individuos de diferentes capacidades sensoriales y metabólicas en el paisaje “Slope”.

### 5.4.4 Estrategia diferente, misma capacidad sensorial y metabólica

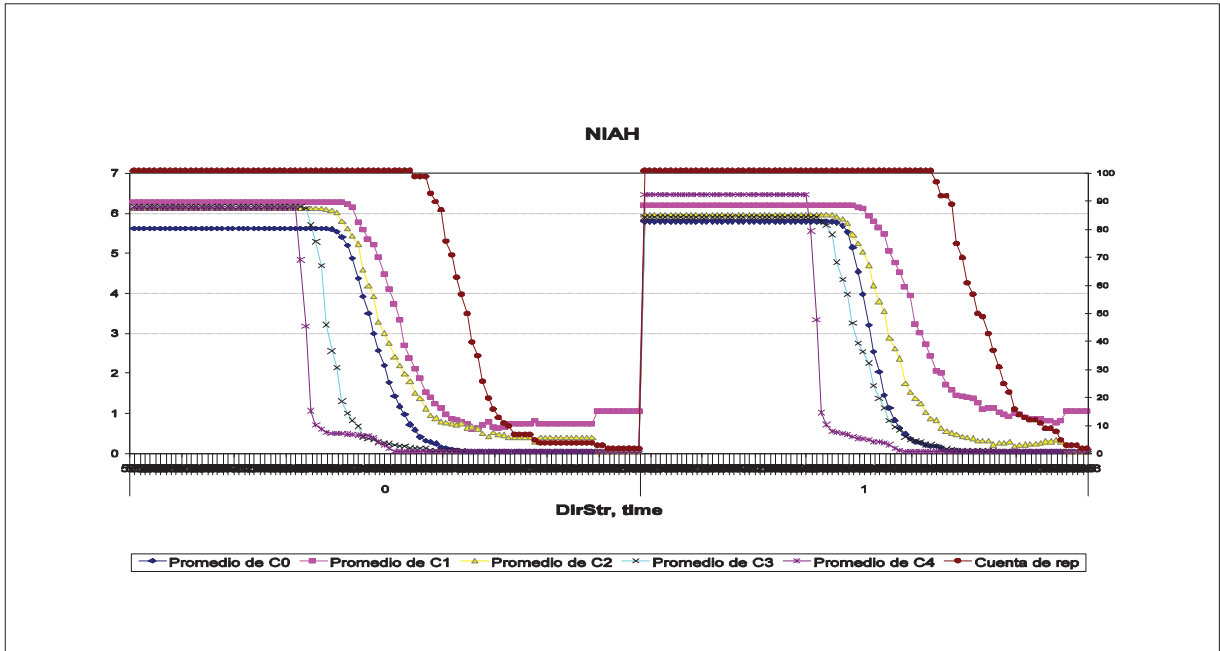
Ahora se verá cuál estrategia tiene mayor robustez en presencia de las otras bajo las mismas condiciones de capacidad sensorial (2), de metabolismo (0.15) y con la misma estrategia de dirección de movimiento.

En un paisaje Random (**Figura 5-19**) la estrategia que muere primero es la 3 en ambos casos de dirección de movimiento. Esto es porque esta estrategia come pocas veces quedando debilitada para la etapa en la que escasea el alimento además de que en esa etapa tiene alta probabilidad de quedar atorado en zonas sin alimento. En  $DirStr = 0$ , la siguiente en morir es la estrategia 4 que no se queda atorada pero tampoco come muy seguido. Le sigue la estrategia 0 que es un poco menos eficiente que las demás ya que en ocasiones lleva a cabo acciones sin un sentido lógico como hacer dos observaciones seguidas, etc. Esta estrategia es mejor que 3 y 4 porque frecuentemente come. La estrategia 1 tiene una secuencia lógica pero come menos seguido y se mueve más (aleatoriamente) que la estrategia 2 la cual acumula más energía que todas en menos tiempo y por esto, cuando escasea el alimento, tiene una buena reserva de energía sobreviviendo más tiempo.

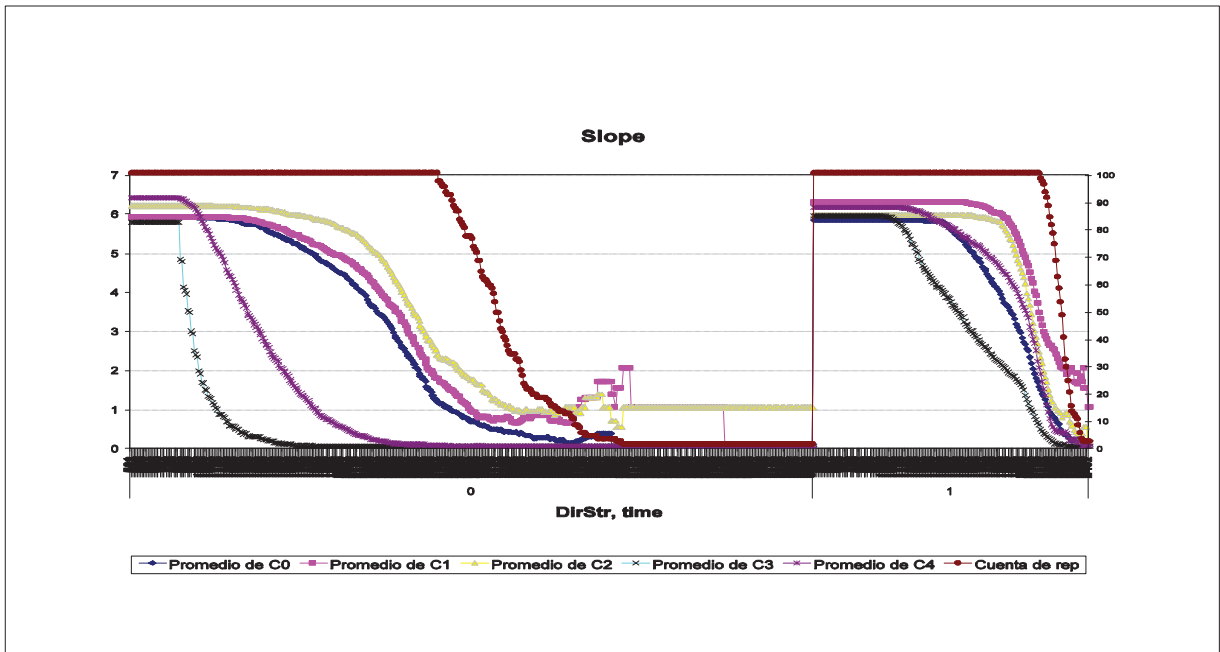
Cuando  $DirStr = 1$  la mejor estrategia ya no es 2 sino 1 porque la estrategia 1 tiene un buen equilibrio y buena lógica en sus comandos y, como ya no hay movimientos aleatorios (ya que  $DirStr = 1$ ) que distraigan la atención hacia zonas que no necesariamente prometen ser buenas, llega más rápidamente a buenas zonas de alimento que la estrategia 2.



**Figura 5-19** Gráfica del promedio de individuos de cada estrategia presentes (escala de la izquierda) y del número de repeticiones que tienen al menos un individuo vivo (escala de la derecha) en función del tiempo para  $DirStr = 0$  (izquierda) y  $DirStr = 1$  (derecha) en el paisaje "Random".



**Figura 5-20** Gráfica del promedio de individuos de cada estrategia presentes (escala de la izquierda) y del número de repeticiones que tienen al menos un individuo vivo (escala de la derecha) en función del tiempo para  $DirStr = 0$  (izquierda) y  $DirStr = 1$  (derecha) en el paisaje “NIAH”.



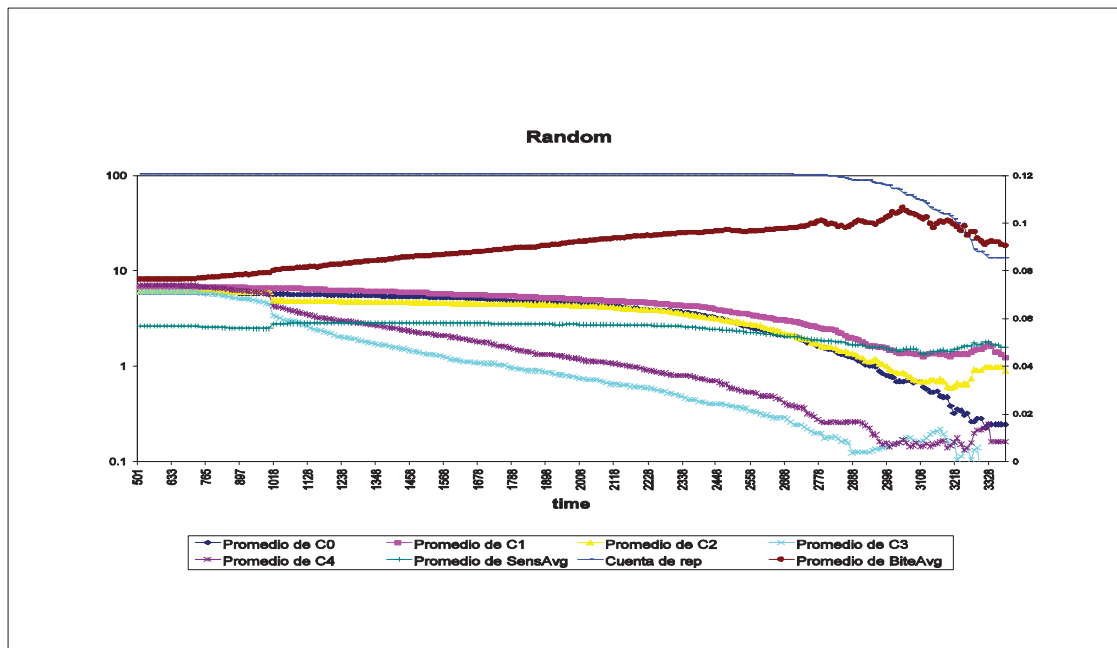
**Figura 5-21** Gráfica del promedio de individuos de cada estrategia presentes (escala de la izquierda) y del número de repeticiones que tienen al menos un individuo vivo (escala de la derecha) en función del tiempo para  $DirStr = 0$  (izquierda) y  $DirStr = 1$  (derecha) en el paisaje “Slope”.

En el caso del paisaje NIAH (**Figura 5-20**) se tiene el mismo comportamiento en ambos casos de DirStr solo que si DirStr = 1 los individuos parecen vivir un poco más de tiempo. La estrategia dominante es en ambos casos la 1 por el buen equilibrio de comandos que ejecuta sin quedarse mucho en un lugar como 2 ni caminar demasiado sin comer como 3 o 4.

Para el caso Slope (**Figura 5-21**) la estrategia dominante depende de DirStr cualitativamente igual que en el caso Random aunque con diferentes valores.

### 5.4.5 Estrategia diferente, diferente capacidad sensorial y metabólica

Finalmente se verá cuál es la configuración dominante de entre todas las posibles (dentro de los rangos establecidos) tratando de averiguar los grados de libertad efectivos en esta situación.



**Figura 5-22** Gráfica del promedio de individuos presentes de cada estrategia de comandos C0, C1, C2, C3 y C4 (escala de la izquierda), del promedio de capacidad sensorial (escala de la izquierda), del número de repeticiones que tienen al menos un individuo vivo (escala de la izquierda) y del promedio de la capacidad metabólica (escala de la derecha) en función del tiempo en el paisaje “Random”.

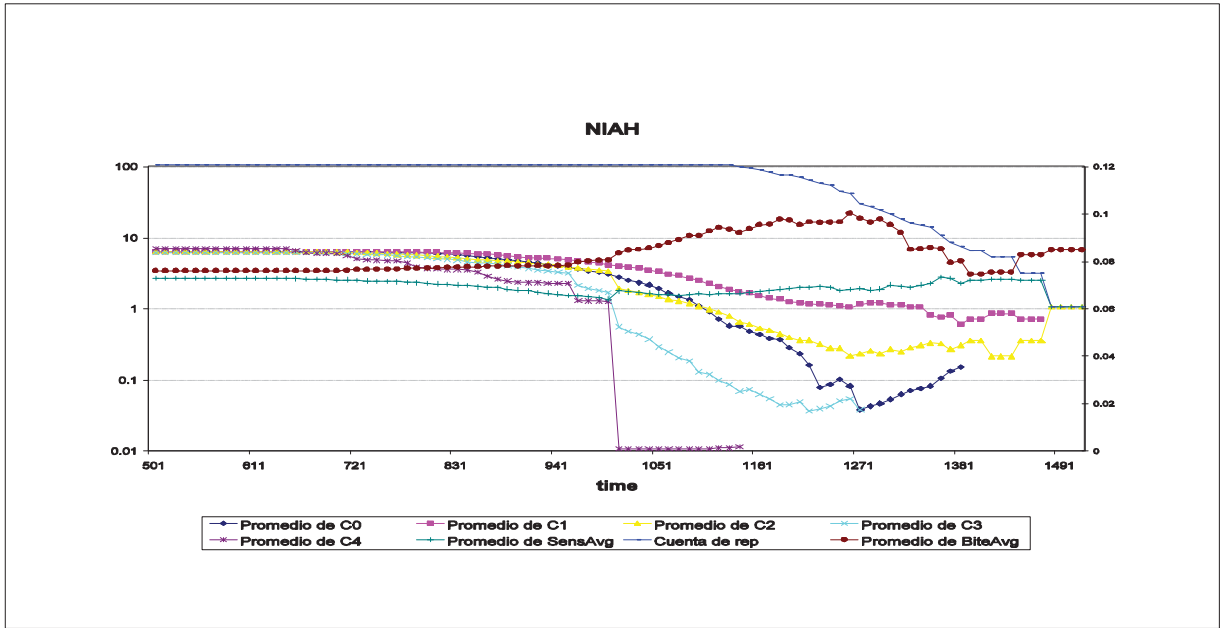


Figura 5-23 Gráfica del promedio de individuos presentes de cada estrategia de comandos C0, C1, C2, C3 y C4 (escala de la izquierda), del promedio de capacidad sensorial (escala de la izquierda), del número de repeticiones que tienen al menos un individuo vivo (escala de la izquierda) y del promedio de la capacidad metabólica (escala de la derecha) en función del tiempo en el paisaje “NIAH”.

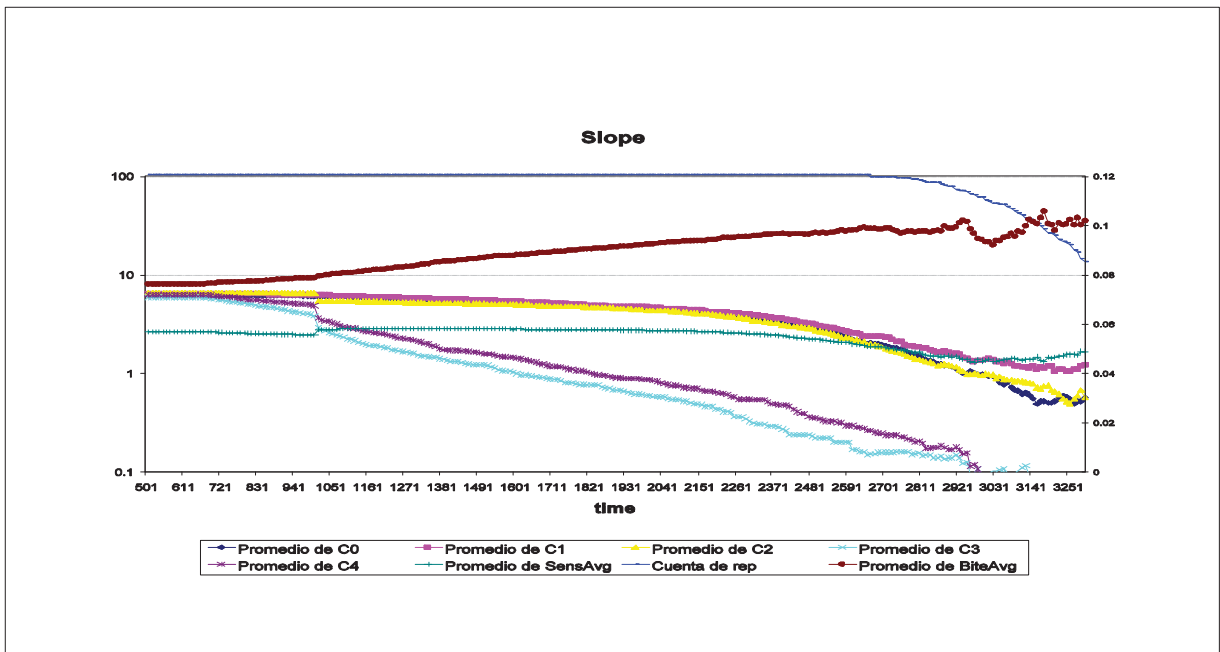


Figura 5-24 Gráfica del promedio de individuos presentes de cada estrategia de comandos C0, C1, C2, C3 y C4 (escala de la izquierda), del promedio de capacidad sensorial (escala de la izquierda), del número de repeticiones que tienen al menos un individuo vivo (escala de la izquierda) y del promedio de la capacidad metabólica (escala de la derecha) en función del tiempo en el paisaje “Slope”.

La mejor estrategia es la 1 (ver **Figura 5-22**, **Figura 5-23** y **Figura 5-24**). El equilibrio entre acciones de consumo y de ingesta es el grado de libertad efectivo emergente en este caso. Se requiere una dedicación a tareas de búsqueda pero no excesiva ya que se debe ingerir alimento periódicamente para mantener una reserva energética adecuada. Por otro lado, si los individuos no hacen búsqueda suficiente e ingieren al máximo, no podrán estar situados cerca de las mejores fuentes de alimento. El mejor sensor es algo un poco mayor a 1 ya que este sensor es relativamente barato en cuanto a consumo de energía pero da la información básica para tomar decisiones. En un porcentaje de las corridas es un sensor de 2 el que domina, de allí que se diga que el mejor es “un poco mayor que 1”. El mejor tamaño de mordida es el más grande pero no es el más seleccionado ya que la selección por capacidad sensorial es más fuerte y, de hecho, los individuos son seleccionados casi independientemente del tamaño de la mordida lo que hace que en las gráficas se tenga una mordida de alrededor de 0.1 y no 0.15 que es la máxima cantidad en los experimentos.

## 5.5 Conclusiones

En el modelo estudiado en este capítulo se da una coevolución ya que la aptitud de un individuo depende de lo que hacen los demás. Esto hace que las características fenotípicas preferidas por el proceso evolutivo sean distintas que cuando los individuos se desarrollan independientemente uno del otro. Se vio que no siempre los rasgos fenotípicos se desarrollan al máximo a favor del individuo. La evolución de rasgos fenotípicos se ve influenciada por el comportamiento de la población como grupo. Esto implica que la evolución no solamente toma en cuenta la aptitud de los individuos (que es importante) para determinar qué rasgos fenotípicos serán favorecidos sino también la aptitud del grupo en su conjunto para hacer sobrevivir a la especie bajo las condiciones presentes en el medio. En este sentido, los grados de libertad efectivos son la capacidad individual para competir con otras especies por recursos y la capacidad de los individuos de cooperar con la conservación de su especie incluso cuando esto significa un sacrificio individual. En el caso estudiado en este capítulo, se ha visto como emergen 3 niveles de grados de libertad efectivos al ir introduciendo cada vez mayor complejidad: 1) la eliminación de competidores de manera que los recursos alcancen mejor para unos pocos que serán los que mantengan viva a una cierta especie, 2) la eficiencia en el uso de la energía desarrollando capacidades adecuadas a las condiciones presentes y 3) el equilibrio en la toma de decisiones de manera que permita mantener un buen nivel energético y a su vez hacer uso de las capacidades para incrementar la probabilidad de hallar alimento.

Este sistema puede hacerse mucho más complejo introduciendo más variables. Ese nivel de complejidad superior se deja para estudios posteriores. Algunas líneas de investigación abiertas son: Introducir capacidad (o aptitud) reproductiva en los individuos de tal forma que puedan transmitir a sus descendientes sus propias características fenotípicas, regeneración no uniforme del alimento para estudiar rasgos como migración por ejemplo, introducir un genoma que codifique los rasgos fenotípicos y una tasa de mutación al momento de la reproducción así como entrecruzamiento de individuos para identificar la manera en que los operadores genéticos logran la adaptación de las especies identificando los grados de libertad efectivos estudiados en los capítulos anteriores, introducir especies presa-predador y estudiar

la evolución de comportamientos de defensa y ataque en función de las poblaciones existentes de dichas especies así como, por lo que se vio en este capítulo, la evolución en los individuos de incapacidades en lugar de más capacidades que permiten sobrevivir mejor a la especie. En fin, el sistema se puede seguir desarrollando mucho todavía y se puede llegar a aprender mucho del análisis detallado de los grados de libertad efectivos emergentes al ir agregando de forma sistemática más complejidad.



## Notas

## *Capítulo 6*      **Conclusiones Finales**

En esta tesis se hace un estudio muy detallado de los grados de libertad efectivos presentes en un algoritmo genético simple (SGA) que evoluciona mediante selección (proporcional a la aptitud), mutación y recombinación (de varias formas) tomando algunos elementos de la literatura disponible pero también se hacen contribuciones originales encaminadas a entender el porqué y el cómo funcionan este tipo de sistemas. Se va de los casos más simples hacia los más complejos.

Se estudian detalladamente los grados de libertad efectivos de cada operador genético por separado y los que emergen de las interacciones entre ellos bajo diferentes condiciones que cubren los casos básicos y los no tan básicos, cosa que no se había hecho anteriormente. El lector debería quedarse con una buena base para inferir lo que pasaría en casos distintos a los estudiados basándose en el comportamiento observado y explicado. Esto es importante para muchos tipos de lectores como son: los teóricos de la computación evolutiva, teóricos en biología molecular y genética, diseñadores de algoritmos evolutivos, usuarios de algoritmos evolutivos, etc.

Se destaca que, a partir del conocimiento generado, fue posible identificar y “corregir” algunas de las limitaciones insalvables del algoritmo genético simple. Esto fue posible mediante modificaciones simples a los operadores genéticos lográndose una clase de algoritmos que resultan mucho más eficientes y eficaces en la búsqueda genética. Al mismo tiempo, se obtuvo una nueva perspectiva para entender el funcionamiento de la (co)evolución biológica y una filosofía aplicable también en otras áreas de la computación evolutiva y adaptativa.

Al identificarse los grados de libertad efectivos en este tipo de sistemas adaptativos, se pueden estudiar sistemas más complejos como el que se describe en el Capítulo 4 o el del Capítulo 5, por ejemplo, e identificar cuales rasgos del sistema tienen su origen en los grados de libertad efectivos presentes en ellos por ser sistemas evolutivos (como los que se estudian en este trabajo) y cuales provienen de las características propias del sistema estudiado. Mucho hay por hacer en el estudio de sistemas evolutivos y adaptativos ya que estos son generalmente muy complejos. Con el presente trabajo se ha dado un paso hacia la comprensión de los fundamentos teóricos que explican el cómo funcionan los sistemas evolutivos complejos tanto naturales como artificiales.

Hay algunas otras contribuciones prácticas de valor en esta tesis como son: la definición de las Curvas Características de Búsqueda (página 195) que resultan una métrica universal para el análisis de diferentes tipos de algoritmos de búsqueda; los operadores genéticos dependientes del ranqueo; la generación de un simulador de vida artificial que puede ser usado para futuros estudios sobre sistemas con mayor complejidad.

# Apéndice 1: Simetrías en la Integración de la Ecuación de Evolución

La ecuación (1.1) representa a un sistema de  $2^N$  ecuaciones con  $2^N$  variables de estado que son las  $P_i(t)$ . Del lado derecho de cada ecuación hay  $2^N$  términos. El esfuerzo computacional para integrar numéricamente todas las ecuaciones es  $O(2^N 2^N) = O(4^N)$  en cada generación lo cual es muy limitante en el valor de  $N$ . Este esfuerzo puede reducirse significativamente cuando las cantidades  $P_i(t)$  son las mismas para varias  $i$  distintas. Este es el caso cuando el paisaje de aptitud está dado mediante un mismo valor para todas las secuencias pertenecientes a una misma clase de Hamming (las que tienen un mismo número de diferencias con alguna secuencia de referencia) y cuando además la población inicial de cada secuencia perteneciente a una misma clase es la misma.

Así, si se define a  $P_{[k]}$  como la suma de las concentraciones de las secuencias pertenecientes a la clase de Hamming  $k$ , la ecuación (1.1) se escribe como:

$$P_{[k]}(t+1) = \sum_{j=1}^{2^N} P_{[m]}(t) p^{d_{i,j}} (1-p)^{N-d_{i,j}} \frac{f_{[m]}(t)}{f(t)}, \quad i \in [k], \quad j \in [m]$$

donde

$d_{i,j}$  es la distancia de Hamming entre una secuencia fija  $i \in [k]$  y la secuencia  $j$ ,

$f_{[m]}$  es la aptitud de la clase a la que pertenece la secuencia  $j$ ,

$P_{[m]}$  es la concentración de las secuencias que pertenecen a la misma clase que  $j$ .

Con lo anterior se reduce el esfuerzo de  $O(2^N 2^N)$  a  $O(N 2^N)$ . Es posible reducir aún más este esfuerzo si se considera que algunos términos estarán repetidos del lado derecho de cada ecuación. Los términos se repiten cuando tienen la misma  $d_{i,j}$  y la misma  $m$ . Es decir, dada una  $i$  fija, existe un número de  $j$ 's cuya distancia de Hamming con  $i$  es la misma y hay un número de ellas que su clase de Hamming es la misma.

Considerando este hecho, la ecuación se transforma en:

$$P_{g^k}(t+1) = \sum_{g=0}^N C_k^g \sum_{h=|k-g|}^{N-|N-k-g|} p^g (1-p)^{N-g} \frac{f_{gh}(t)}{f(t)} \frac{P_{gh}(t)}{C_h^g} C_{(g+h-k)/2}^{N-k} C_{(g+h-k)/2}^g$$

donde la segunda sumatoria es sobre valores para  $h$  con un espaciamiento de 2.

Esta nueva ecuación se integra con un esfuerzo de  $O(NN^2) = O(N^3)$  por generación lo cual significa una reducción muy importante que permite hacer integraciones con valores de  $N$  mucho mayores o para un número mucho mayor de generaciones.

## Apéndice 2: Curvas Características de Búsqueda

Se definen aquí las Curvas Características de Búsqueda las cuales ofrecen un panorama global del desempeño de un algoritmo específico sobre un problema dado.

Estas curvas se obtienen evaluando, para un número grande de repeticiones (se recomienda usar unas 100 repeticiones o más), el esfuerzo (número de evaluaciones de la función de aptitud) que cada repetición requirió para encontrar el óptimo, luego se ordenan las repeticiones en orden ascendente desde las que usaron menos esfuerzo hasta las que usaron más y se grafica este esfuerzo contra las repeticiones ordenadas obteniéndose así una curva monótona creciente.

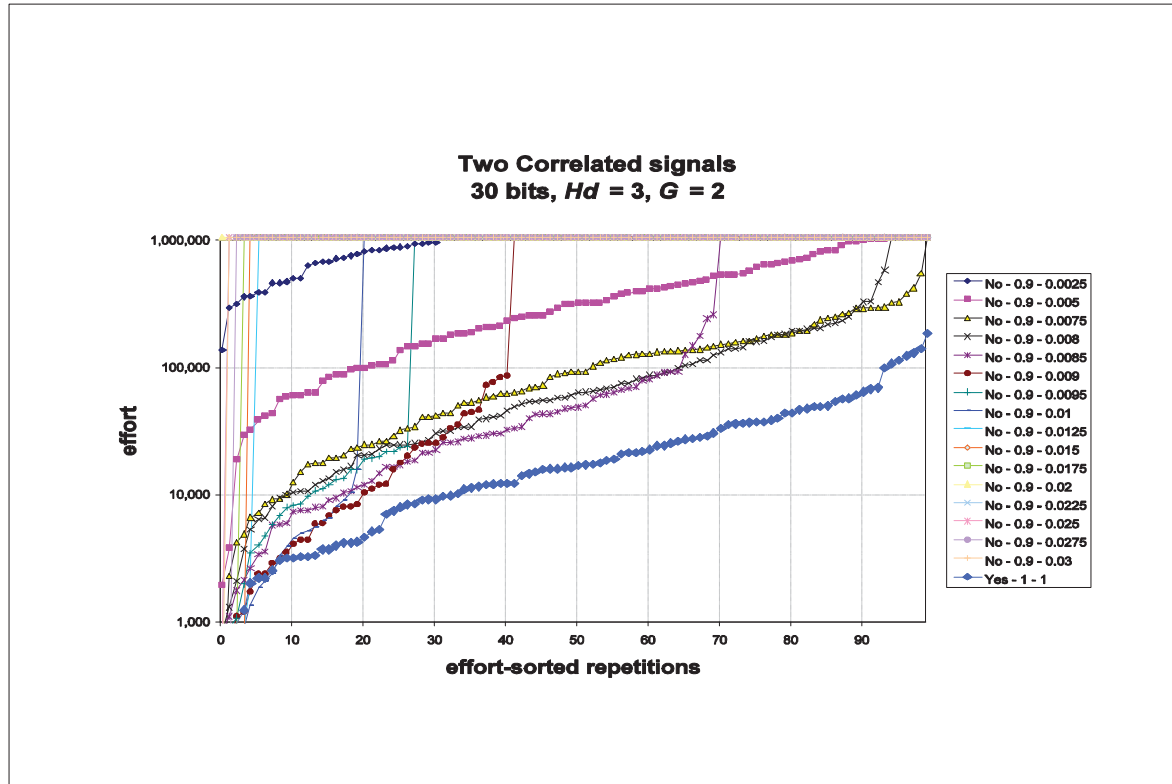
Con estas curvas es muy fácil obtener mucha información acerca del comportamiento del algoritmo de búsqueda como son: esfuerzo promedio (aproximado), esfuerzo máximo, esfuerzo mínimo, moda (aproximada), etc. Se puede ver además, a simple vista, si hay algún fenómeno que afecte solamente a una fracción de las repeticiones y qué porcentaje de éstas se ve afectado. Es fácil ver también qué porcentaje de las repeticiones son exitosas, para hallar el óptimo, con un presupuesto fijo de esfuerzo computacional y, además, se puede ver cuánto beneficio da el aumentar ese presupuesto.

Estas curvas presentan ruido, el cual disminuye si se aumenta el número de repeticiones y entonces las curvas son cada vez más suaves mostrando con más precisión el comportamiento esperado del algoritmo.

Un ejemplo típico de estas curvas se muestra en la **Ilustración 1** de este apéndice. En ella se han puesto varias Curvas Características de Búsqueda asociadas cada una con una variante distinta de algoritmos de búsqueda sobre un mismo problema.

La manera de leerlas es la siguiente: Iniciando por el eje vertical que indica el esfuerzo, se toma un valor fijo y se observa el punto correspondiente de la curva para ese valor y se obtiene, en el eje horizontal, la fracción de las repeticiones que pueden hallar el óptimo con el esfuerzo fijado o menor. Así, las curvas más bajas corresponden siempre a un mejor desempeño del algoritmo que las más altas. Sin embargo, en ocasiones como en el caso del ejemplo, algunas curvas se llegan a cruzar. Esto se debe a que en una fracción de las repeticiones se da algún fenómeno distinto que hace que en esas repeticiones el algoritmo tenga mayor o menor dificultad para hallar el óptimo. En la gráfica del ejemplo se ve que algunas curvas se doblan hacia arriba indicando la presencia de algún fenómeno solo en una fracción de las repeticiones. Todo esto se distingue a simple vista notándose claramente en qué porcentaje de las repeticiones se da. También es posible ver rápidamente el valor máximo y mínimo de esfuerzo que cada algoritmo puede tener además de que es claro si estos valores

máximo y mínimo fueron resultado de buena/mala suerte o si son valores comunes. En el caso del valor mínimo puede verse si el algoritmo puede resolver el problema con poco esfuerzo de manera consistente y en el caso del máximo si el algoritmo tiene dificultad en muchas ocasiones o no.

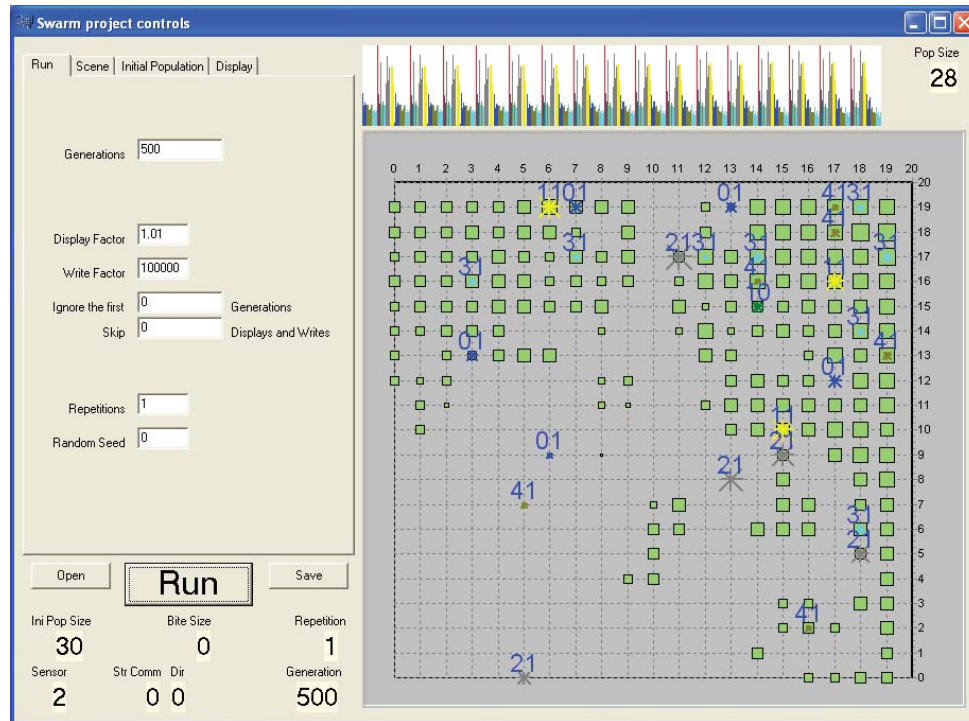


**Ilustración 1** Ejemplo de comparación de las Curvas Características de Búsqueda correspondientes a diferentes versiones de algoritmos de búsqueda sobre un mismo problema.

Estas curvas pueden entonces ser usadas como una *métrica universal* para evaluar y comparar el desempeño de distintos algoritmos resolviendo un mismo problema y/o el desempeño de un algoritmo sobre diferentes problemas.

## Apéndice 3: Opciones del simulador

En pantalla pueden hacerse cambios en la configuración de la simulación de la siguiente manera.



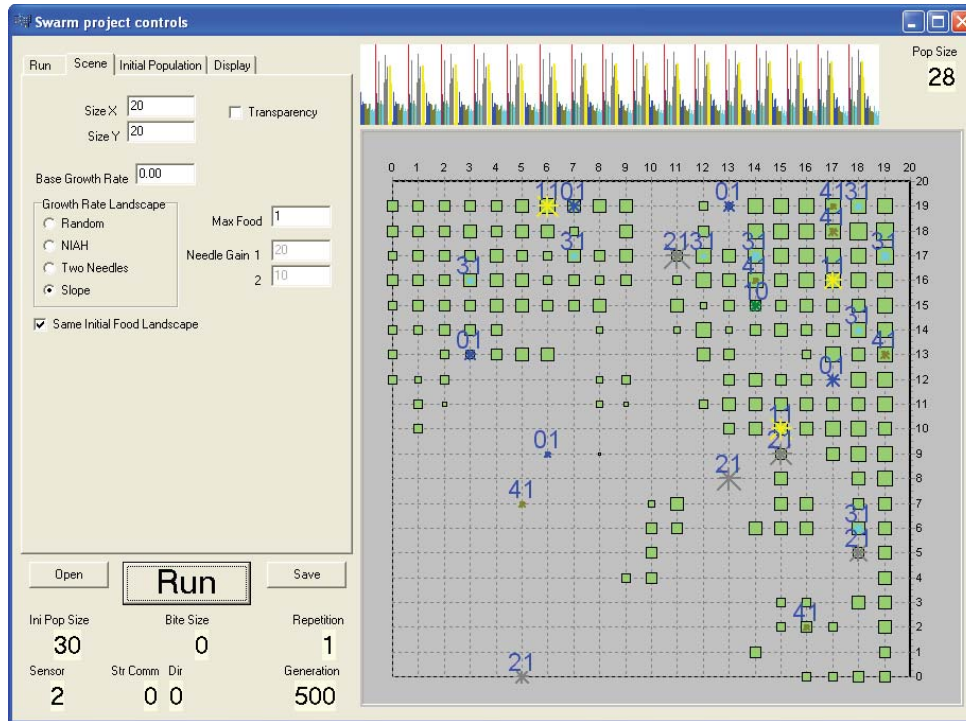
**Ilustración 2** Pestaña “Run”.

En la pestaña “Run” (Ilustración 2) se fijan:

- ? *Generations* indica el número de lapsos de tiempo o generaciones máximo de las corridas.
- ? *Display Factor* es el factor de muestra en pantalla. Este factor indica en que generación se actualizará la información del panel de la simulación. Esto es para hacer que las simulaciones muestren cada una de las generaciones o que se salten algunas para acelerarlas. La siguiente generación que será mostrada es la que supera al valor de la última generación mostrada multiplicada por este factor.
- ? *Write Factor* es el factor de escritura en disco. Este factor funciona igual que el anterior pero controla la generación en la que se enviarán los datos de la corrida al archivo en disco.
- ? *Ignore the first \_\_\_\_ generations.* El número de generaciones que inicialmente no se deberán mostrar en pantalla ni escribir en disco.
- ? *Skip \_\_\_\_ Displays and Writes.* El número de muestras en pantalla y de escrituras en disco

iniciales que no se harán.

- ? *Repetitions*. El número de repeticiones que deberán hacer con los parámetros actuales.
- ? *Random Seed*. La semilla del generador de números aleatorios. Cuando es cero, se inicializa con un valor que depende de la fecha y hora del día actual.



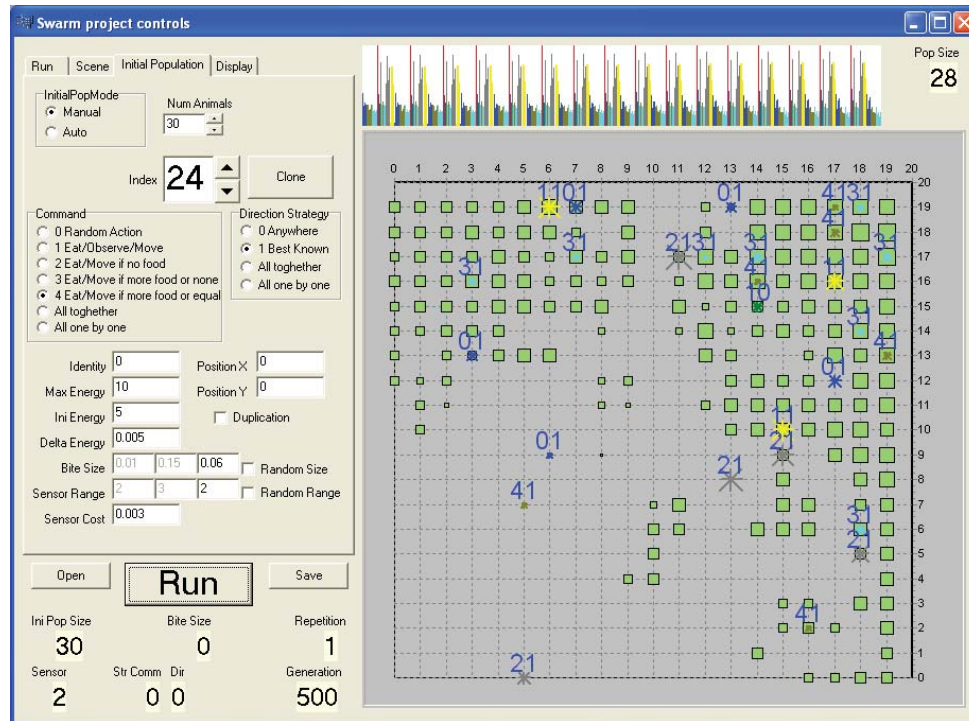
**Ilustración 3** Pestaña “Scene”.

En la pestaña “Scene” (Ilustración 3). Se fijan:

- ? El tamaño del terreno en número de columnas y de renglones de casillas.
- ? La tasa base de crecimiento del alimento. Esta tasa base se usa en conjunto con el paisaje de tasas de crecimiento.
- ? El paisaje de tasas de crecimiento. Se dan 4 opciones:
  - o *Random* asigna a cada casilla una tasa de crecimiento aleatoria entre 0 y la tasa base (ver punto anterior).
  - o *NIAH* asigna la tasa base a una sola casilla y a las demás les asigna la tasa base dividida entre la ganancia de la aguja 1 (ver abajo).
  - o *Two Needles* asigna a una casilla la tasa base y a otra la tasa base entre la ganancia de la aguja 1 por la ganancia de la aguja 2 y al resto le asigna la misma tasa que *NIAH*.
  - o *Slope* (o lineal) asigna una tasa de cero a la casilla en el origen y a las demás una tasa proporcional a su distancia al origen haciendo que la tasa más alta sea la tasa base.
- ? *Max Food* es la cantidad máxima de alimento que puede haber en una casilla. En caso de exceder esta cantidad, simplemente se recorta el valor a ésta.
- ? *Needle Gain 1* es la ganancia de tasa de crecimiento de la aguja respecto a la paja.



- ? *Needle Gain 2* es la ganancia de tasa de crecimiento de la segunda aguja.
- ? *Same Initial Food landscape* indica si se usará el mismo paisaje para inicializar las cantidades de alimento en las casillas o si se inicializará en forma aleatoria. Si se sigue el mismo paisaje entonces se usa el esquema de inicialización de las tasas de crecimiento pero para las cantidades iniciales.



**Ilustración 4** Pestaña “Initial Population”.

En la pestaña “Initial Population” (Ilustración 4) se fijan:

- ? El método de inicialización de los individuos. *Manual* o *Auto*.
  - o En modo *Manual* se inhibe la captura de los parámetros *Num Animals Max* y *Num Animals Min*, así como *Bite Size Min*, *Bite Size Max*, *Sensor Range Min*, *Sensor Range Max*, estos cuatro últimos se inhiben solamente si *Random Size* está apagado. *Bite Size Aux* y *Sensor Range Aux* se inhiben solamente si *Random Range* está encendido.
  - o En modo *Auto* se inhibe el botón *Clone* y los campos *Index*, *Identity*, *Position X* y *Position Y*.
- ? *Num Animals Min* determina el número inicial de animales en la simulación.
  - o En modo *Auto* este número cambia automáticamente desde *Num Animals Min* hasta *Num Animals Max* incrementándose *Num Animals Incr* en cada ocasión y se hacen todas las repeticiones para cada uno de los valores.
- ? *Num Animals Max*, en modo automático, determina el valor máximo del número inicial de animales.
- ? *Num Animals Incr* determina el incremento en el número inicial de animales para la siguiente serie de repeticiones.

- ? *Index* indica el índice del animal cuyos parámetros están siendo editados.
- ? Botón *Clone* hace una copia de los parámetros del individuo que está siendo editado en el siguiente individuo.
- ? *Command* especifica el tipo de estrategia de los individuos. En modo *Auto* es la estrategia general. En modo *Manual* es la estrategia del individuo identificado con el campo *Index*. Sus opciones son:
  - o *Random Action*. Se elige una acción al azar.
  - o *Eat/Observe/Move*. Se hace una rotación en ese orden de las acciones en cada generación.
  - o *Eat/Move if no food*. Se hace una observación y si ya no hay suficiente alimento se ordena la acción *move* de lo contrario la acción es *Eat*.
  - o *Eat/Move if more food or none*. Se hace una observación y si ya no hay suficiente alimento o se detecta una casilla con más alimento que la actual se ordena la acción *move* de lo contrario la acción es *Eat*.
  - o *Eat/Move if more food or equal*. Se hace una observación y si ya no hay suficiente alimento o se detecta una casilla con más o igual cantidad de alimento que la actual se ordena la acción *move* de lo contrario la acción es *Eat*.
  - o *All together*. Se distribuyen todas las anteriores en la población uniformemente.
  - o *All one by one*. Se hacen todas las repeticiones para cada uno de las estrategias posibles.
- ? *Direction Strategy* indica la forma en que será decidida la dirección del movimiento en los individuos. En modo *Auto* es para toda la población y en modo *Manual* para el individuo *Index*. Sus opciones son:
  - o *Anywhere*. La dirección del movimiento es aleatoria.
  - o *Best Known*. Hacia la casilla con más alimento.
  - o *All together*. Se distribuyen en la población ambas opciones aleatoriamente.
  - o *All one by one*. Se hacen las repeticiones con cada una de las opciones.
- ? *Identity*. Es un identificador del individuo. Puede o no usarse. Es solo para marcar algunos individuos en particular.
- ? *Max Energy*. Nivel máximo de energía que puede alcanzar el individuo *Index (Manual)* o toda la población (*Auto*).
- ? *Ini Energy*. Nivel de energía inicial del individuo *Index (Manual)* o toda la población (*Auto*).
- ? *Delta Energy*. Costo mínimo de energía en cada generación del individuo *Index (Manual)* o toda la población (*Auto*).
- ? *Position X, Position Y*. Posición inicial del individuo *Index (Manual)* o toda la población (*Auto*).
- ? *Duplication*. Indica si el individuo *Index (Manual)* o toda la población (*Auto*) podrá(n) reproducirse por clonación una vez que alcance(n) el nivel *Max Energy* convirtiéndose (cada individuo) en 2 individuos idénticos pero con la mitad de la energía.
- ? *Bite Size Min, Bite Size Max, Bite Size Aux*.
  - o En modo *Manual*, y si *Random Size* está activada, el tamaño de la mordida del individuo *Index* será inicializado con un valor entre *Bite Size Min* y *Bite Size Max*.
  - o En modo *Manual* y con *Random Size* apagada, el individuo *Index* tendrá inicialmente una mordida de tamaño *Bite Size Aux*.
  - o En modo *Auto* y si *Random Size* está activada, se harán todas las corridas con cada

uno de los valores desde *Bite Size Min* hasta *Bite Size Max* con incrementos de *Bite Size Aux*, que indican el tamaño máximo de mordida inicial la cual será aleatoria entre 0 y valor máximo.

- En modo *Auto* y si *Random Size* está apagada, se harán todas las corridas con cada uno de los valores desde *Bite Size Min* hasta *Bite Size Max* con incrementos de *Bite Size Aux*, que indican el tamaño máximo de mordida inicial en todos los individuos.
- ? *Sensor Range Min, Sensor Range Max, Sensor Range Aux.*
- En modo *Manual*, y si *Random Range* está activada, el alcance del sensor del individuo *Index* será inicializado con un valor entre *Sensor Range Min* y *Sensor Range Max*.
  - En modo *Manual* y con *Random Range* apagada, el individuo *Index* tendrá inicialmente un alcance de sensor de *Sensor Range Aux*.
  - En modo *Auto* y si *Random Range* está activada, se harán todas las corridas con cada uno de los valores desde *Sensor Range Min* hasta *Sensor Range Max* con incrementos de *Sensor Range Aux*, que indican el alcance máximo de sensor el cual será aleatorio entre 0 y valor máximo.
  - En modo *Auto* y si *Random Range* está apagada, se harán todas las corridas con cada uno de los valores desde *Sensor Range Min* hasta *Sensor Range Max* con incrementos de *Sensor Range Aux*, que indican el alcance máximo de sensor en todos los individuos.
- ? *Sensor Cost.* Indica el costo en energía que sufren los individuos por cada dato obtenido con el sensor. Los sensores de mayor alcance tendrán más costos.

En la pestaña “Display” (Ilustración 5) se indica:

- ? *Show Range.* Muestra en pantalla el alcance de cada individuo.
- ? *Show Strategy.* Muestra en pantalla la estrategia de cada individuo.
- ? *Show Identity.* Muestra en pantalla la identidad de cada individuo.

*Open.* Abre un archivo con datos previamente guardados con el botón *Save*.

*Run.* Inicia los experimentos.

*Save.* Guarda todos los datos capturados en las pestañas para posteriormente poder repetir experimentos o crear variantes de los mismos.

Se muestra en pantalla el terreno con cada uno de los individuos. Las diferentes especies se identifican con un color distinto cada una.

En la parte superior del terreno se muestra el nivel de energía de cada individuo presente en la población. Junto a esta grafica se tiene el número de individuos presentes.

En la parte inferior se muestran: El tamaño inicial de la población, el tamaño de la mordida (máximo o exacto), el número de repetición de la corrida, la capacidad sensorial, la estrategia de comandos, la dirección de movimiento y el número de generación en la repetición en curso.

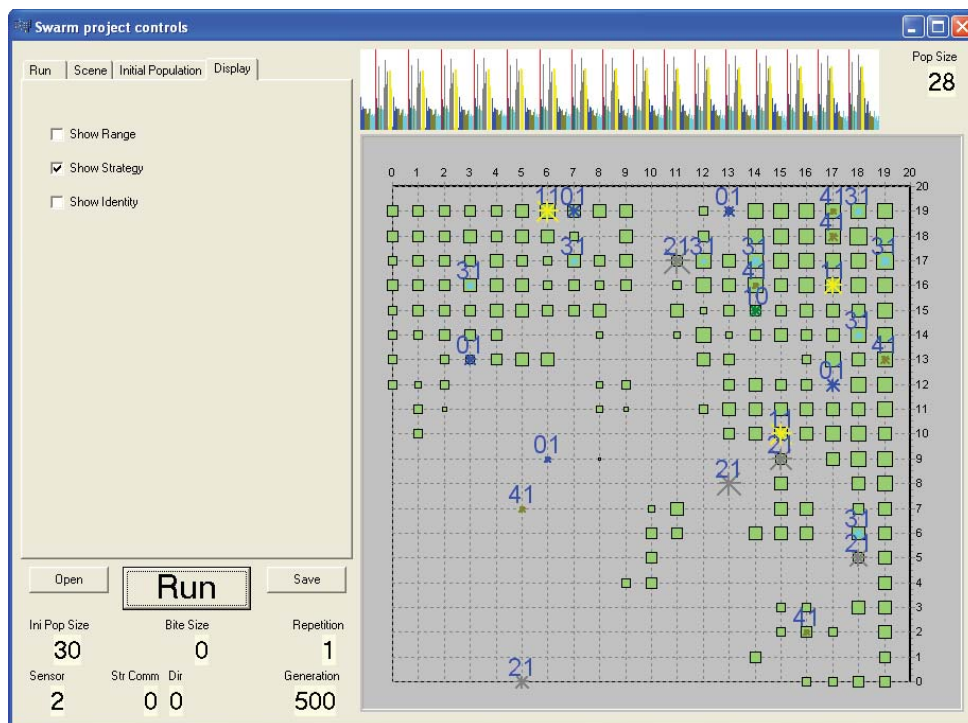


Ilustración 5 Pestaña “Display”.



_5526_7561_7346_7706_7656_4646_2270_2430	__+hx_0+xh_0h+x_00hx_0x+x_+x+x_hh0h_h+hh	'+x+x+x+x0	100
_5526_7561_7346_7706_7656_4646_2270_2430	__+hx_0+xh_0h+x_00hx_0x+x_+x+x_hh0h_h+hh	'+x+x+x+x0	100
_5526_7561_7346_7706_7656_4646_2270_2430	__+hx_0+xh_0h+x_00hx_0x+x_+x+x_hh0h_h+hh	'+x+x+x+x0	100
_5526_7561_7346_7706_7656_4646_2270_2430	__+hx_0+xh_0h+x_00hx_0x+x_+x+x_hh0h_h+hh	'+x+x+x+x0	100
_5526_7561_7346_7706_7656_4646_2270_2430	__+hx_0+xh_0h+x_00hx_0x+x_+x+x_hh0h_h+hh	'+x+x+x+x0	100
_5526_7561_7346_7706_7656_4646_2270_2430	__+hx_0+xh_0h+x_00hx_0x+x_+x+x_hh0h_h+hh	'+x+x+x+x0	100
_5526_0136_0416_1645_7454_4646_2750_2410	__+hx_hhhx_h+h_x_hx++_0+++_+x+x_h0+h_h+hh	'+x+x+x+xh+h_xh0	23.7769
_5526_0136_0416_1645_7454_4646_2750_2410	__+hx_hhhx_h+h_x_hx++_0+++_+x+x_h0+h_h+hh	'+x+x+x+xh+h_xh0	23.7769
_5526_0136_0416_1645_7454_4646_2750_2410	__+hx_hhhx_h+h_x_hx++_0+++_+x+x_h0+h_h+hh	'+x+x+x+xh+h_xh0	23.7769
_5526_0136_0416_1645_7454_4646_2750_2410	__+hx_hhhx_h+h_x_hx++_0+++_+x+x_h0+h_h+hh	'+x+x+x+xh+h_xh0	23.7769
_5526_0136_0416_1645_7454_4646_2750_2410	__+hx_hhhx_h+h_x_hx++_0+++_+x+x_h0+h_h+hh	'+x+x+x+xh+h_xh0	23.7769
_5526_0136_0416_1645_7454_4646_2750_2410	__+hx_hhhx_h+h_x_hx++_0+++_+x+x_h0+h_h+hh	'+x+x+x+xh+h_xh0	23.7769
_5526_0136_0416_1645_7454_4646_2750_2410	__+hx_hhhx_h+h_x_hx++_0+++_+x+x_h0+h_h+hh	'+x+x+x+xh+h_xh0	23.7769
_5526_0136_0416_1645_7454_4646_2750_2410	__+hx_hhhx_h+h_x_hx++_0+++_+x+x_h0+h_h+hh	'+x+x+x+xh+h_xh0	23.7769
_5526_0136_0416_1645_7454_4646_2750_2410	__+hx_hhhx_h+h_x_hx++_0+++_+x+x_h0+h_h+hh	'+x+x+x+xh+h_xh0	23.7769
_5506_7561_2422_1341_5654_4646_2660_2430	__+hx_hhhx_h+h_x_hx++_0+++_+x+x_h0+h_h+hh	'+x+x+x+xh+h_xh0	23.7769
_5526_0210_1646_6656_6704_5461_3764_3376	__+hx_hhhx_h+h_x_hx++_0+++_+x+x_h0+h_h+hh	'+x+x+x+xh+h_xh0	23.7769
_6670_6203_5547_7706_7705_6363_4746_2201	__+hx_hhhx_h+h_x_hx++_0+++_+x+x_h0+h_h+hh	'+x+x+x+xh+h_xh0	23.7769
_7511_7254_3477_4631_3255_6642_0270_5657	__+hx_hhhx_h+h_x_hx++_0+++_+x+x_h0+h_h+hh	'+x+x+x+xh+h_xh0	23.7769
_5601_4762_5265_4434_1644_4663_4237_1637	__+hx_hhhx_h+h_x_hx++_0+++_+x+x_h0+h_h+hh	'+x+x+x+xh+h_xh0	23.7769
_2777_7071_5632_6106_7614_0147_2067_6312	__+hx_hhhx_h+h_x_hx++_0+++_+x+x_h0+h_h+hh	'+x+x+x+xh+h_xh0	23.7769
_4623_1654_1607_2636_6406_2735_1622_6207	__+hx_hhhx_h+h_x_hx++_0+++_+x+x_h0+h_h+hh	'+x+x+x+xh+h_xh0	23.7769
_3773_6526_6461_6347_5275_7102_7764_1656	__+hx_hhhx_h+h_x_hx++_0+++_+x+x_h0+h_h+hh	'+x+x+x+xh+h_xh0	23.7769
_1627_6755_1062_1644_5550_3713_4367_5705	__+hx_hhhx_h+h_x_hx++_0+++_+x+x_h0+h_h+hh	'+x+x+x+xh+h_xh0	23.7769
_5261_6762_7231_3360_6115_6365_3646_6211	__+hx_hhhx_h+h_x_hx++_0+++_+x+x_h0+h_h+hh	'+x+x+x+xh+h_xh0	23.7769
_4041_3717_0026_4545_6533_2626_7271_6514	__+hx_hhhx_h+h_x_hx++_0+++_+x+x_h0+h_h+hh	'+x+x+x+xh+h_xh0	23.7769
_4051_3757_0026_4545_6537_2226_7271_6510	__+hx_hhhx_h+h_x_hx++_0+++_+x+x_h0+h_h+hh	'+x+x+x+xh+h_xh0	23.7769
_7551_7255_2465_0264_4672_6646_0744_6405	__+hx_hhhx_h+h_x_hx++_0+++_+x+x_h0+h_h+hh	'+x+x+x+xh+h_xh0	23.7769
_4132_5744_4477_5641_6604_5463_6177_4540	__+hx_hhhx_h+h_x_hx++_0+++_+x+x_h0+h_h+hh	'+x+x+x+xh+h_xh0	23.7769
_4133_0433_7067_7501_6672_3200_6004_1202	__+hx_hhhx_h+h_x_hx++_0+++_+x+x_h0+h_h+hh	'+x+x+x+xh+h_xh0	23.7769
_6545_6546_1341_6437_1742_0664_6445_0616	__+hx_hhhx_h+h_x_hx++_0+++_+x+x_h0+h_h+hh	'+x+x+x+xh+h_xh0	23.7769
_5740_1716_1223_6026_3436_6647_6250_6507	__+hx_hhhx_h+h_x_hx++_0+++_+x+x_h0+h_h+hh	'+x+x+x+xh+h_xh0	23.7769
_2113_7267_6672_1641_7654_4636_3265_1513	__+hx_hhhx_h+h_x_hx++_0+++_+x+x_h0+h_h+hh	'+x+x+x+xh+h_xh0	23.7769
_5733_7141_0372_3235_7705_4677_7104_1472	__+hx_hhhx_h+h_x_hx++_0+++_+x+x_h0+h_h+hh	'+x+x+x+xh+h_xh0	23.7769
_5630_4520_6103_1353_3657_4746_7324_7637	__+hx_hhhx_h+h_x_hx++_0+++_+x+x_h0+h_h+hh	'+x+x+x+xh+h_xh0	23.7769
_5526_7561_6261_1641_7654_4346_2631_2430	__+hx_hhhx_h+h_x_hx++_0+++_+x+x_h0+h_h+hh	'+x+x+x+xh+h_xh0	23.7769
_7756_5253_6477_3414_6733_7050_5171_5006	__+hx_hhhx_h+h_x_hx++_0+++_+x+x_h0+h_h+hh	'+x+x+x+xh+h_xh0	23.7769



Emergencia de Grados de Libertad Efectivos en Sistemas Adaptativos

	x0hh_0h+h_+h0h_+hhx		
	__+xh_+0hh_hhxx_hh0x_		
_4462_4702_2262_3176_4336_3765_3274_1476	+hhx_h0x+_hh0+_h+0x	'+hhx+hhxhh0	1.97483
	_h++0_xhx0_hxhh_h+h0_		
_2557_6267_1632_1427_6614_7665_5047_5117	xxh+_0xx+_h+0_+hh0	'hx	1.89872
	_+hhh_0h++_x+hh_h+0h_		
_5231_7255_6433_3470_2662_3667_4310_1737	hxhx_hxx0_+hhh_h0h0	'hx	1.89872
	_+x+h_++0h_+hh+_hhhh_		
_5653_5572_5235_1013_0714_2626_6646_2473	h0h+_hxhx_xx+x_h+0h	'hx	1.89872
	_h+hh_0+xh_hhhh_h+x0_		
_3500_7561_0022_2567_0041_7467_0650_1657	hh+h_0+x0_hx+h_hx+0	'h+x0	1.89872
	_+0xh_x0xh_hxxh_hhxx_		
_5761_6760_0662_0266_5764_3653_6631_4040	+0x+_hx+h_xxhh_+h+h	'hx	1.89872
	_h0h0_h+hh_h+xh_h++h_		
_2727_2432_3562_0442_1356_5264_2422_0027	hh+x_+hx+_h+hh_hhh0	'h+xhhhh0	1.89872
	_+0h+_0+hx_+hxh_x0hh_		
_5715_7516_5261_6703_6654_0642_7105_3765	xx++_hx+h_0hh+_h0x+	'hx	1.89872
	_++h+_0+x+_xhx0_+hh+_		
_5524_7565_6267_4215_7650_1674_3231_3637	0x+h_hx0+_hhhh_hxh0	'hx	1.89872
	_+hx_0+xh_h+hh_hhhh_		
_5526_7561_0422_2001_3121_0426_2717_2630	hhhh_h+hx_h0h0_hxhh	'h+hxh+hxh+hhh0h0	1.89318
	_hh00_+h+h_xh+h_+hhh_		
_1377_4041_6242_4200_6665_7063_3637_1673	xxx+_0hx_hxh0_hx0h	'+h+h+hhhhx0hhx0	1.88112
	_h++0_0hh+_hhhh_xhxx_		
_0547_7015_2122_6226_5306_3735_6573_6706	+hhx_h0h+_x+0h_x0hx	'h++0h0h++hhxx0	1.86857
	_+hhh_hx00_hh+h_hhhx_		
_5111_0677_0042_1126_4365_1166_1601_7305	+hx+_hhxx_hxhh_0hh+	'hhx	1.86736
	_+0hh_hhhx_x+xh_hhhh_		
_5733_0006_6562_1122_6335_2166_1601_5677	xhh+_hhxx_hxhh_+x00	'hhx	1.86736
	_+h++_xhhx_+hh+_h+hh_		
_4254_6106_4324_1412_3262_6410_1340_7102	hhxh_x+hh_hh+h_0hhh	'hhx	1.86736
	_+00x_+hh_hhhh_0hhh_		
_5776_5432_2030_7300_6511_2161_3631_1750	x+hh_hhxh_hxhh_h0+h	'hhx	1.86736
	_hhhh_x0+h_h+xx_hhxh_		
_3231_6741_3466_0263_5765_4262_3121_3300	+0x+_+hxh_hhhh_hhhh	'hhx	1.86736
	_hhxh_hxhh_0hxh_+hh_		
_0360_1632_7263_4531_3744_0526_2713_7330	h0++_h+hx_h0hh_0hhh	'hhx	1.86736
	_+xhh_0+xh_hxhh_+x+h_		
_5621_7561_0622_5641_7654_1366_6310_5637	0x++_hhxx_xhhh_+xh0	'hhx	1.86736
	_+hx_0+xh_h+hh_+0hx_		
_5526_7561_0422_4726_2406_2262_2670_6430	h+hx_hhxh_hx0h_x+hh	'hhx	1.86736
	_+0hh_+h0+_xxhh_xxhh_		
_4722_5275_6603_6600_0257_3012_1662_2657	hh+0_hhhh_hxxh_hx+0	'hh+0hx	1.85178
	_hhhh_hhhx_h+hh_hx+h_		
_0112_2316_3423_1641_5654_4647_2770_1720	+x++_+x+0_h00h_h0hh	'hhhhhhhx	1.82566
	_0h0+_0hx0_+hh+_hh0h_		
_7274_7267_4335_0173_7666_7533_5124_7215	0xxx_0+hh_+hh+_0hh+	'0	1.76998
	_hxxx_+hh0_hh++_xhhx_		
_1666_4207_0055_6226_0726_1520_7434_3210	h0hx_h+h_h_0+h+_hhhh	'+hh00	1.76998
	_0xxh_hxh+_hhh+_x0hh_		
_7661_0614_1204_6703_7447_2364_4105_3765	0++0_hhx+_+hh+_h0x+	'h0	1.76998
	_0hhx_+x+h_+hh+_+hh_		
_7116_5642_5215_4531_6715_5627_6327_7523	x0h+_+xh0_xhh0_0+hh	'0	1.76998
	_0+h+_h+0h_hhhh_hhx+_		
_7425_2471_3133_1365_5722_2772_4135_4230	+0hh_h00h_+hh+_+hhh	'+hhh+0hhhhhh00	1.76998
	_xhhh_hhh+_hhh+_+h0x_		
_6132_0014_0205_5276_7270_4343_7731_2614	0h0h_+h+h_00hh_hxh+	'0	1.76998
	_+x0_x+0h_hhhx_hhxh_		
_4567_6470_2206_0161_1177_4622_3601_1350	hh00_+xhh_hxhh_hh+h	'hh0	1.76998
	_+xxh_xx0h_0xh+_h+0h_		
_5661_6672_7604_2472_2755_5776_7264_7573	h0++_+00x_0hx+_0+0h	'+00	1.76998
	_+hhh_hxh+_h+hh_hx+h_		
_4123_0634_0522_3641_7654_4646_7315_0762	0x++_+x+x_0hh+_h0xh	'0	1.76998
	_+x0_+h+h_hh+h_h+++_		
_4567_5352_1350_1444_7654_5246_7633_5300	0x++_+h+x_0xhh_+hhh	'0	1.76998
	_0x0h_x0+h_hx+h_x0hx_		
_7673_6742_3642_6706_7224_5632_2242_2272	0hh+_+xhh_hh+h_hh0h	'hh0	1.76998
	_0++x_h+h0_00+0_00hh_		
_7546_2417_7747_7720_6437_5204_0372_3710	x+h0_+hh+_hh0h_h0hh	'h0	1.76998
	_hh0x_0+xh_h+hh_hh+h_		
_0376_7561_0422_1241_7654_4646_2770_6430	0x++_+x+x_h00h_x+hh	'hh0	1.76998

\_1214\_3073\_6001\_7765\_7654\_4662\_2670\_2430    \_hh+\_hh0h\_xhhh\_00x+\_  
0x+\_+xxh\_hx0h\_h+hh    'hh0    1.76998



## Referencias

- [ 1 ] Aguirre. *Grados de libertad efectivos en algoritmos genéticos*. Master Thesis UNAM 1999.
- [ 2 ] Altenberg, L. (1995). The Schema Theorem and Price's Theorem. In: *Foundations of Genetic Algorithms 3* (Darrell Whitley and Michael D. Vose. Eds.). Morgan Kaufmann. Estes Park, Colorado, USA. pp. 23-49.
- [ 3 ] Angeles, O., Waelbroeck, H. Stephens, C.R. (1998) Emergence of Algorithmic Language in Genetic Systems, *BioSystems* 47, pp. 129-147.
- [ 4 ] Bäck, Optimal Mutation Rates in Genetic Search. In *Proceedings of ICGA 5*, ed. S. Forrest, 2-8, Morgan Kaufmann (1993).
- [ 5 ] Bäck, Self-adaptation in Genetic Algorithms, In *Proceedings of 1st European Conference on Artificial Life*, 263-271 MIT Press (1991).
- [ 6 ] Bäck, T. (1996). *Evolutionary Algorithms in Theory and Practice*. Oxford University Press.
- [ 7 ] Banzhaf, W., P. Nordin, R.E. Keller y F.D. Francone (1998) *Genetic Programming: An Introduction*. Morgan Kaufmann Publishers.
- [ 8 ] Beyer, Schwefel. (2002) *Evolution Strategies. A comprehensive introduction*. *Natural Computing* 1: 3-52, 2002. Kluwer Academic Publishers. Netherlands.
- [ 9 ] Borland C++ Builder Enterprise Suite Version 6.0 © 1983-2002 Borland Software Corporation.
- [ 10 ] Beyer. An alternative explanation for the manner in which genetic algorithms operate. *BioSystems*, 41(1):1-15, 1997.
- [ 11 ] Bürger, R. (2000). *The Mathematical Theory of Selection, Mutation and Recombination* Wiley, Chichester, Inglaterra.
- [ 12 ] Cervantes, Stephens. "Optimal" mutation rates for genetic search. *Proceedings of Genetic and Evolutionary Computation Conference (GECCO) '06*, pp. 1313-1320, Maarten Keijzer et. al. (2006).
- [ 13 ] Cervantes, Stephens. A Rank Proportional Generic Genetic Algorithm. *Lecture Series on Computer and Computational Sciences*, Volume 7, 2006, pp 71-74. Brill Academic

Publishers.

- [ 14 ] Cervantes, Stephens. Limitations of existing mutation rate heuristics and how a Rank GA overcomes them. *IEEE Transactions on Evolutionary Computation*. In review.
- [ 15 ] Chung, S.W. and R.A. Perez (1994). The Schema Theorem considered insufficient. In *Proceedings of the Sixth IEEE International Conference on Tools with Artificial Intelligence*. New Orleans. pp. 748-751.
- [ 16 ] Coello. Van Veldhuizen, Lamont. *Evolutionary Algorithms for Solving Multi-Objective Problems*. Springer 2002. ISBN 0306467623.
- [ 17 ] Drossel, B., *Biological Evolution and Statistical Physics*, a ser publicado en *Advances in Physics*, preprint cond-mat/0101409 (2001).
- [ 18 ] Eigen and Schuster. *The Hypercycle: A Principal of Natural Self-Organization*, Springer-Verlag (1979).
- [ 19 ] Eigen, McCaskill, Schuster. *The Molecular Quasi-species*. *Adv. Chem. Phys.* 75, 149, 1989.
- [ 20 ] Fogarty, Varying the Probability of Mutation in the Genetic Algorithm, *ICGA 3*, ed. J.D. Schaffer, 104-109, Morgan-Kaufmann (1989).
- [ 21 ] Fogel, D.B. and A. Ghozeil (1997). Schema Processing under Proportional Selection in the Presence of Random Effects. *IEEE Transactions on Evolutionary Computation* 1 (3), 269-288.
- [ 22 ] Fogel, D.B. and A. Ghozeil (1998). The Schema Theorem and the misallocation of trials in the Presence of Stochastic Effects. In: *Evolutionary Programming VII: Proc. Of the 7th Ann. Conf. On Evolutionary Programming*
- [ 23 ] Forrest and Mitchell. Relative building-block fitness and the building-block hypothesis. In L. Darrell Whitley, editor, *Foundations of genetic algorithms 2*, pages 109–126, San Mateo, 1993. Morgan Kaufmann.
- [ 24 ] Geiringer, H. (1944) On the Probability Theory of Linkage in Mendelian Heredity, *Annals of Mathematical Statistics* 15, pp. 25-27.
- [ 25 ] Goldberg, 1989. *Genetic algorithms in search, optimization and machine learning*. Addison Wesley, Reading, Massachusetts.
- [ 26 ] Grefenstette, Optimization of Control Parameters for Genetic Algorithms, *IEEE Trans SMC*, 16(1), 122-128 (1986).
- [ 27 ] Grefenstette. Deception considered harmful. In *Foundations of Genetic Algorithms*,

- pages 75–91, San Mateo, 1992. Morgan Kaufmann.
- [ 28 ] Harik and Goldberg D.G. Learning linkage. In FOGA 4, pages 247–262, 1996.
  - [ 29 ] Hesser and Männer, Towards an Optimal Mutation Probability for Genetic Algorithms, LNCS 496, ed. H.P. Schwefel, Springer Verlag (1991).
  - [ 30 ] Holland, 1975. Adaptation in natural and artificial systems. MIT Press, Cambridge, Massachusetts.
  - [ 31 ] Ishibuchi and Narukawa. Recombination of similar parents in EMO algorithms. Lecture Notes in Computer Science 3410:Evolutionary Multi-Criterion Optimization, pp.265-279, Springer, Berlin, March 2005.
  - [ 32 ] Ishibuchi and Shibata. A Similarity-Based Mating Scheme for Evolutionary Multiobjective Optimization. E.Cantú-Paz et al.(Eds.):GECCO 2003, LNCS 2723, pp.1065 –1076,2003. © Springer Verlag Berlin Heidelberg 2003.
  - [ 33 ] Jansen, Wegener. Real royal road functions where crossover provably is essential. Discrete Applied Mathematics, 149(1-3):111–125, 2005.
  - [ 34 ] Jones. Evolutionary Algorithms, Fitness Landscapes and Search. PhD. Thesis 1995.
  - [ 35 ] Kauffman S (1993). The Origins of Order. Self Organization and Selection in Evolution. Oxford University Press. USA.
  - [ 36 ] Kimura (1983). The neutral theory of molecular evolution. Cambridge University Press.
  - [ 37 ] Koza, J.R. (1992). Genetic Programming: On the Programming of Computers by Means of Natural Selection. MIT Press.
  - [ 38 ] Langdon, W.B. y Poli, R. (2002) Foundations of Genetic Programming Springer Verlag, Alemania.
  - [ 39 ] Maple 7, Waterloo Maple Inc. Copyright © 1981-2001.
  - [ 40 ] Mitchell, Forrest, and Holland. The royal road for genetic algorithms: Fitness landscapes and GA performance. In F. J. Varela and P. Bourguine, editors, Proceedings of the First European Conference on Artificial Life, pages 243–254, Cambridge, MA, 1992. MIT Press.
  - [ 41 ] Mühlenbein, How Genetic Algorithms Really Work: Mutation and Hill Climbing, PPSN II, editors B. Männer and R. Manderick, 15-25, North Holland (1992).
  - [ 42 ] Nowak, Komarova 2001. Towards an evolutionary theory of language. Trends in

cognitive science Vol 5. No 7.

- [ 43 ] Ochoa, Error Thresholds in Genetic Algorithms. Evolutionary Computation, MIT Press (in press).
- [ 44 ] Ochoa, Harvey and Buxton, Error Thresholds and their Relation to Optimal Mutation Rates. In D. Floreano, J-D. Nicoud, and F. Mondada (Eds.) ECAL'99, Lecture Notes in Artificial Intelligence 1674, pp 54-63, Springer-Verlag, Berlin (1999).
- [ 45 ] Ochoa, Harvey and Buxton, On Recombination and Optimal Mutation Rates.
- [ 46 ] Ochoa, Setting the Mutation Rate: Scope and Limitations of the 1/L Heuristics, GECCO-2002, pp 315-322, Morgan Kaufmann, San Francisco, CA (2002).
- [ 47 ] Poli R., "Schema theory without expectations for GP and GAs with one-point crossover in the presence of schema creation," in *Foundations of Genetic Programming* (T. Haynes, W. B. Langdon, U.-M. O'Reilly, R. Poli, and J. Rosca, eds.), (Orlando, Florida, USA), 13 July 1999.
- [ 48 ] Poli, R. (2000). Hyperschema Theory for GP with one-point crossover, building blocks and some new results in GA theory, In R. Poli, W. Banzhaf, W.B. Langdon, J. Miller, P. Nordin and T.C. Fogarty, Editors, Proceedings of EuroGP2000, pp. 163-180, Springer-Verlag, Berlin.
- [ 49 ] Poli, R. and Langdon W.B. (1997). A New Schema Theory for Genetic Programming with One-Point Crossover and Point Mutation, in Proc. of Second Annual Conference on Genetic Programming, (J.R. Koza, D.E. Goldberg, D.B. Fogel,
- [ 50 ] Poli, R. y Freitag N. (2001). Exact Schema Theory for GP and Variable-Length GAs with Homologous Crossover, in Proc. of GECCO 2001, Morgan Kaufmann CA.
- [ 51 ] Poli, R., J.E. Rowe y N. Freitag (2001). Markov Models for GP and Variable-Length GAs with Homologous Crossover, in Proc. of GECCO 2001, Morgan Kaufmann CA.
- [ 52 ] Prugel Bennett, A. and Shapiro, J.L. (1994) Phys. Rev. Lett. 72, pp 1305-1308.
- [ 53 ] Reidys, C.M. y Stadler, P.F. (2002). Combinatorial Landscapes, SIAM Review 44, pp. 3-54.
- [ 54 ] Sewell, Samarabandu, Rodrigo, and McIsaac. The Rank-scaled Mutation Rate for Genetic Algorithms. International Journal of Information Technology, Volume 3 Number 1, 2006.
- [ 55 ] Smith and Fogarty, Self Adaptation of Mutation Rates in a Steady State Genetic Algorithm, Proceedings of CEC, 318 - 323, IEEE Press (1996).

- [ 56 ] Stephens (2001). Some Exact Results from a Coarse Grained Formulation of Genetic Dynamics, in Proc. of GECCO 2001, Morgan Kaufmann CA.
- [ 57 ] Stephens (2002). The renormalization group and the dynamics of genetic systems. Acta physica slovacica vol. 52 no. 6,515-524.
- [ 58 ] Stephens, Cervantes. Just what are Building Blocks? FOGA 2007.
- [ 59 ] Stephens, Garcia Olmedo, Mora Vargas, Waelbroeck, Self-Adaptation in Evolving Systems, Artificial Life, Vol. 4, Issue 2, 183-201 (1998).
- [ 60 ] Stephens, Mora Vargas. Effective Fitness as an Alternative Paradigm for Evolutionary Computation I: General Formalism, Genetic Programming and Evolvable Machines Volume 1, 363-378 (2000).
- [ 61 ] Stephens, Zamora, Wright. Perturbation Theory and Renormalization Group in Genetic Dynamics. FOGA 2005.
- [ 62 ] Stephens, Waelbroeck 1997. Effective degrees of freedom in genetic algorithms and the block hypothesis. In Bäck, T., editor, Proceedings of the Seventh International Conference on Genetic Algorithms, pages 34-41, Morgan Kaufmann, San Mateo, California.
- [ 63 ] Stephens, Waelbroeck 1998. Analysis of effective degrees of freedom in genetic algorithms. Physical Review, D57:3251-3264.
- [ 64 ] Stephens, Waelbroeck 1999. Schemata evolution and building blocks. Evolutionary Computation 7(2):109-124 MIT press.
- [ 65 ] Stephens, Waelbroeck, Aguirre. (1999), Schemata as Building Blocks: Does Size Matter?, Foundations of Genetic Algorithms 5, ed. W. Banzhaf (Morgan Kaufmann, San Mateo)
- [ 66 ] Stephens, Waelbroeck, Angeles. (1998), Emergence on Algorithmic Language in Genetic Systems. Biosystems 47, 129-147.
- [ 67 ] Stephens. (2005) Zero is not a four letter word. Euro GP 2005. Lecture Notes in Computer Science. Springer Berlin/Heidelberg Volume 3447/2005 pp. 371-380.
- [ 68 ] Vose, M.D. (1991). Generalizing the Notion of Schema in Genetic Algorithms, Artificial Intelligence 50(3), pp. 385-396.
- [ 69 ] Vose, M.D. and Liepins, G. (1991). Punctuated Equilibria in Genetic Search, Complex Systems 5, pp. 31-44.
- [ 70 ] Vose, M.D. and Nix, A. (1992). Modelling Genetic Algorithms with Markov Chains,

Annals of Mathematics and Artificial Intelligence 5, pp. 79-98.

- [ 71 ] Vose, M.D. The simple genetic algorithm: Foundations and theory. MIT Press, Cambridge, MA, 1999.
- [ 72 ] Watson. Analysis of recombinative algorithms on a non-separable building-block problem. In Foundations of Genetic Algorithms, pages 69–89, San Mateo,2001. Morgan Kaufmann.
- [ 73 ] Wright, S. (1932). The roles of mutation, in breeding, crossbreeding and selection in evolution, Proceedings of the Sixth International Congress on Genetics, ed. D.F. Jones, pp 356-366.

# Notas