

03063



UNIVERSIDAD NACIONAL
AVENIDA DE
MEXICO

**UNIVERSIDAD NACIONAL AUTÓNOMA
DE MÉXICO**

POSGRADO EN CIENCIA E INGENIERÍA DE LA COMPUTACIÓN

**“CAMBIO DE ESCALA Y DINÁMICA DE GRADOS
DE LIBERTAD EFECTIVOS EN SISTEMAS
GENÉTICOS ARTIFICIALES”**

T E S I S

QUE PARA OBTENER EL GRADO DE:

**MAESTRO EN CIENCIAS
(COMPUTACIÓN)**

P R E S E N T A:

ANDRÉS AGUILAR CONTRERAS

DIRECTOR DE TESIS:

DR. CHRISTOPHER RHODES STEPHENS STEVENS

México, D.F.

2004.



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.

Resumen

En el presente trabajo hacemos una revisión de las bases biológicas de la evolución y las comparamos con sus contrapartes artificiales, seguimos las derivaciones de las ecuaciones de evolución, examinamos el concepto de cambio de escala y mostramos su relevancia en algunos escenarios donde aparecen simetrías, resolvemos las ecuaciones de evolución para los escenarios típicos planteados y pocos grados de libertad, hacemos comentarios sobre como estos modelos de juguete pueden darnos intuición del comportamiento en escenarios más generales.

CAMBIO DE ESCALA Y DINÁMICA DE
GRADOS DE LIBERTAD EFECTIVOS EN
SISTEMAS GENÉTICOS ARTIFICIALES

Andrés Aguilar Contreras

2004

*A mi madre,
que con tanto trabajo y esfuerzo
me ha impulsado a salir adelante.
Siempre has sido un ejemplo*

Índice general

1. Introducción	2
2. Preliminares	4
2.1. Evolución	4
2.1.1. Evolución natural	4
2.1.2. Evolución artificial	7
2.2. Aptitud y escenarios de aptitud (espacio de estados/configuraciones)	9
2.3. Dinámica	14
2.4. Cambio de escala	18
2.5. Algoritmos genéticos	19
2.5.1. Cuerdas	20
2.5.2. Representación	21
2.5.3. Operadores genéticos	22
2.6. Antecedentes sobre la evolución artificial	25
2.7. Cadenas de Markov	28
3. Dinámica de cuerdas	30
3.1. Solo selección	30
3.2. Selección y mutación	32
3.3. Selección y cruza	33
3.4. Representación de “bloques constructores”	38
3.5. Ecuación general microscópica	38
4. Cambio de escala y dinámica de grados de libertad efectivos	41
4.1. Elementos generales	42
4.2. Operadores genéticos escalados	44
4.3. Esquemas	45
4.3.1. Solo selección	46
4.3.2. Selección y mutación	46
4.3.3. Selección y cruza	49
4.3.4. Ecuación general (selección, mutación y cruza)	51
4.4. Mapeo genotipo-fenotipo	51
4.4.1. Escenario plano	52
4.4.2. Contando unos	52

4.4.3. Modelo de Eigen	53
5. Solución de las ecuaciones para pocos grados de libertad en algunos escenarios típicos	56
5.1. Soluciones para selección 1 bit	56
5.1.1. Escenario plano	57
5.1.2. Escenario de contando 1's	57
5.1.3. Modelo de Eigen	58
5.2. Soluciones para selección 2 bits	59
5.2.1. Escenario plano	60
5.2.2. Escenario de contando 1's	60
5.2.3. Modelo de Eigen	61
5.3. Soluciones para selección 3 bits	62
5.3.1. Escenario plano	62
5.3.2. Escenario de contando 1's	62
5.3.3. Modelo de Eigen	63
5.4. Soluciones para selección y mutación 1 bit	64
5.4.1. Escenario plano	65
5.4.2. Contando 1's	65
5.4.3. Modelo de Eigen	67
5.5. Soluciones para selección y mutación 2 bits	69
5.5.1. Escenario plano	69
5.5.2. Escenario de contando 1's	70
5.5.3. Modelo de Eigen	74
5.6. Soluciones para Selección y cruza 2 bits	76
5.6.1. Escenario plano	76
5.6.2. Escenario de contando 1's	77
5.6.3. Modelo de Eigen	78
5.7. Soluciones para Selección y cruza 3 bits	78
5.7.1. Escenario plano	78
6. Soluciones analíticas, soluciones numéricas	80
6.1. Ecuación general formal	80
6.2. Solución para la ecuación de selección	81
6.3. Solución para la ecuación de mutación	84
6.4. Solución para la ecuación de cruza	86
6.5. Comportamiento de la aptitud como función del tiempo considerando solo selección	87
6.6. Análisis de valores propios y vectores propios para selección y mutación	88
7. Conclusiones	91
7.1. Reconocimientos	92

A. Soluciones alternativas y aproximaciones tomando límite de tiempo continuo	93
A.1. Soluciones para selección y mutación 1 bit escenario de contando 1's	93
A.2. Soluciones para selección y mutación 1 bit modelo de Eigen	94
A.3. Soluciones para selección y mutación 2 bits modelo de Eigen	95
A.4. Soluciones para selección y cruza 2 bits escenario plano	96
A.5. Soluciones para selección y cruza 3 bits escenario plano	97
A.6. Derivación para la ecuación de mutación 3 bits contando 1's	99
B. Coarse-graining in genetic algorithms: some issues and examples	101

Capítulo 1

Introducción

La computación evolutiva (CE) es un paradigma nuevo y de rápido crecimiento en el campo de la inteligencia artificial (IA). Se basa en los principios de la evolución natural, es decir, la supervivencia del más apto. Imitando estos procesos naturales pretende resolver problemas que resultan difíciles para los métodos tradicionales en muy distintas áreas como son: la optimización, la clasificación, el aprendizaje y búsqueda en espacios complejos.

Han surgido diferentes ramas de investigación dentro de la CE; cada una pretende explotar diferentes aspectos de esta mímica con la naturaleza, pero al mismo tiempo comparten las ideas centrales. Las ramas que destacan son¹:

- Algoritmos Genéticos (AGs). Éstos manipulan típicamente cadenas binarias de longitud fija y podemos mencionarlos como los de mayor impacto por ser de propósito general.
- Programación Genética (PG). Trabajan sobre programas por ejemplo en el lenguaje en LISP ya que ahí pueden ser expresados mediante estructuras de “árboles de análisis”. Su objetivo es la producción automática de programas de cómputo.
- Estrategias Evolutivas (EE). En ellas se consideran estructuras de datos más variadas como cadenas de longitud variable o variables de tipo real, y se permiten algunas variantes, por ejemplo en el tamaño de la población.
- Programación Evolutiva (PE). Atribuida a Lawrence Fogel, trabaja sobre máquinas de estados.

Los AGs han tenido un gran impacto como estrategia de búsqueda de soluciones sobre espacios complejos, fomentando así su puesta en práctica en muchos sistemas en años recientes. Sus aplicaciones van desde resolver problemas difíciles de optimización hasta aproximar soluciones de problemas que pertenecen a la categoría de NP. Sin embargo, a pesar de su éxito práctico, relativamente poco

¹Estas son las de mayor desarrollo, aunque hay muchas variantes

se ha desarrollado en el aspecto teórico, la gente que se dedica a las aplicaciones asume que deben funcionar muchas veces sin preguntarse por qué.

Fue a mediados de los 70s cuando Holland [11] propuso el término de AG, sin embargo, no fue sino hasta los 90s que Michael D. Vose con su libro "The Simple Genetic Algorithm" [30] sentó las bases para el modelado de los AG como sistemas dinámicos. Alden Wright [16], Darrell Whitley y Jonathan Rowe entre otros han extendido este formalismo [20]. Adam Prügel-Bennett y Shapiro modelan la dinámica mediante técnicas de mecánica estadística como en [17, 18]. Christopher Stephens [28] y Lee Altenberg [3] han desarrollado diferentes conclusiones sobre la importancia del teorema del esquema de Holland en los distintos modelos. Riccardo Poli ha hecho aportaciones al área de PG y ha extendido algunos de estos resultados a estructuras de longitud variable [15, 16]. Colin Reeves y Jonathan Rowe presentan una visión más amplia de lo desarrollado en la teoría en [20]. Este trabajo de tesis está centrado en los resultados obtenidos en estos trabajos tomando como punto de partida el desarrollo del Dr. Christopher Stephens y colaboradores [28, 26, 27, 23, 24, 25] encaminado a formalizar la teoría de AGs canónicos a través de una nueva representación.

Vale la pena mencionar que la falta de avances en la parte teórica es principalmente porque cualquier resultado tiene una gran complejidad. Incluso para los casos más sencillos, notamos que es muy difícil o hasta imposible encontrar soluciones analíticas a las ecuaciones que describen su comportamiento.

En la presente tesis daremos un seguimiento a estos resultados teóricos obtenidos tanto en el área de AGs, como en la de Genética Poblacional. Mostramos resultados clásicos, así como sus contrapartes más actuales. Comparamos dichos resultados teóricos con experimentos y simulaciones para verificarlos. Pretendiendo por supuesto que estos resultados teóricos den sustento a la base empírica existente.

El documento esta estructurado como sigue: En el siguiente capítulo presentamos la base biológica, así como terminología y conceptos generales. En el tercer capítulo nos enfocamos al estudio de cuerdas, que son las estructuras fundamentales de los AGs canónicos. En el capítulo cuarto introducimos el concepto de "coarse graining" o cambio de escala y destacamos su importancia en la teoría desarrollada, tomando como caso particular a los esquemas sobre los que profundizamos. Después, en el capítulo 5 presentamos algunas soluciones explícitas para casos sencillos (modelos de juguete), mostrando la complejidad a la que nos enfrentamos así como el uso de las técnicas de cambio de escala. Por último, en el capítulo 6 presentamos las soluciones formales a las ecuaciones. En los apéndices mostramos algunas soluciones alternativas a las ecuaciones de evolución.

Capítulo 2

Preliminares

2.1. Evolución

El estudio de la evolución tiene gran importancia desde el punto de vista biológico dado que todos los fenómenos biológicos se derivan de relaciones evolutivas e interacciones ocurridas en el pasado. Gracias al avance tecnológico en el área de cómputo podemos ahora simular diferentes escenarios dentro de la evolución. Pero no solo eso, sino que también podemos hacer uso de los conceptos evolutivos y aplicarlos para resolver problemas de ingeniería, economía y muchas otras áreas. Esta visión ha dado origen a lo que conocemos como evolución artificial. En las siguientes secciones mostramos algunos conceptos importantes para la biología y sus contrapartes en el estudio de la evolución artificial.

2.1.1. Evolución natural

En la naturaleza el objetivo primordial de todo organismo es la preservación de su material genético o genotipo, este material tiene codificada la información que determina el fenotipo (composición macroscópica, forma física) del individuo. La presente sección se basa principalmente en [9] y [29], tomando en cuenta el trabajo de Sewell Wright [33] y Kauffman [12].

Todos los organismos vivos están constituidos de células, en cada célula hay el mismo conjunto de cromosomas; los cromosomas son cadenas de ADN y sirven como un modelo para el organismo completo. Podemos dividir a los cromosomas en genes, los genes son bloques de ADN que codifican una proteína particular. De manera general podemos decir que una combinación particular de genes codifica un rasgo distintivo (por ejemplo el color de los ojos). Llamamos alelos a los posibles valores para cada rasgo (por ejemplo azul, café). Cada gen tiene su propia posición en el cromosoma a la que llamamos locus. El material genético completo (o conjunto de todas las combinaciones de cromosomas) es llamado genoma, a un conjunto particular de genes es a lo que llamamos genotipo. El genotipo es la base para el fenotipo del organismo, es decir sus características mentales y físicas, como el color de ojos, la estatura, etc. En la naturaleza esta

información (genotipo) se guarda por lo general en un par de cadenas entrelazadas (diploidismo).¹ El diploidismo ha probado ser útil pues permite almacenar, además de los genes que determinan al individuo (genes dominantes), información de las características que no estarán presentes en el fenotipo del individuo (genes recesivos) que probablemente fueron dominantes en otro ambiente. Esto tiene el objetivo de no perderlas por completo, lo que da la oportunidad de (probablemente) explorarlas nuevamente dentro de otras combinaciones.

En la naturaleza se preserva el material genético mediante la reproducción. En este proceso, primero ocurre una recombinación (cruza) del material genético. Es decir, los genes de los padres de alguna manera se mezclan y forman nuevas combinaciones de cromosomas. La nueva descendencia creada puede presentar pequeñas variaciones o mutaciones en el material genético ocasionadas por errores en el proceso de copiado de los genes de los padres. La mutación tiene aspectos positivos y negativos, pues mientras que alguna variabilidad es ventajosa, por ejemplo para prevenir que una enfermedad acabe con la especie, pues al haber diferencias entre los individuos la enfermedad no los afecta de igual forma. Otro ejemplo es la exploración de nuevas posibilidades, el efecto inmediato de muchas combinaciones puede no serlo. Es decir, la mutación puede contribuir a la "carga genética", es decir, perpetuando alelos recesivos inútiles. La carga genética es el conjunto de características genotípicas que afectan el desarrollo de una población.

La evolución, vista desde el punto de vista de Darwin, requiere de un mecanismo que actúe sobre los individuos al cual conocemos como *selección natural*, que consiste en la proliferación de los individuos "más aptos" o los que tienen características que les dan mejores probabilidades de sobrevivir, por lo tanto decimos que la selección se lleva a cabo en el fenotipo aunque los cambios actúen en el genotipo.

A pesar de que Darwin dio la propuesta para la teoría evolutiva, los resultados formales aparecieron con Mendel quien con sus aportaciones dio lugar a lo que hoy conocemos como genética poblacional. La genética poblacional se interesa por las bases genéticas de la evolución. Difiere mucho de la biología en que sus ideas importantes son teóricas más que observacionales o experimentales. Difícilmente podría ser de otra manera. Sus objetos de estudio son principalmente las frecuencias y aptitudes de genotipos en poblaciones naturales. Podemos decir que la evolución es el cambio en las frecuencias de los genotipos a través del tiempo, quizá por sus diferencias en aptitud. Mientras las frecuencias en los fenotipos son fácilmente medibles, su cambio no lo es. La escala de tiempo de cambio de la mayoría de las variantes genéticas que ocurren naturalmente es muy grande, probablemente del orden de decenas de miles de millones de años.² Cambios así de lentos son imposibles de observar directamente. Las diferencias

¹En el caso natural también existe haploidismo, es decir una sola cadena codifica toda la información e incluso – aunque en menor proporción – el poliploidismo, (con tres o más cadenas)

²La excepción son algunos organismos de rápida reproducción con los que normalmente se experimenta, aunque algunos experimentos de laboratorio con RNA muestran cambios en el orden de minutos

de aptitud entre genotipos, que pueden ser responsables de algunos cambios de frecuencias, son tan extraordinariamente pequeñas, probablemente menos del 0.01 por ciento, que también son imposibles de medir directamente. A pesar de que podemos observar directamente el estado de la población, no hay realmente una forma de explorar directamente la evolución de la población. Algún progreso se ha hecho en genética poblacional construyendo modelos matemáticos de la evolución, estudiando su comportamiento.

En los 30's quedó claro que la evolución es un fenómeno de poblaciones que puede ser representado como un conjunto de cambios en las frecuencias de los genes en la población a causa de la acción de diversas fuerzas naturales como selección y "genetic drift" [33]. Estos cambios pueden llevar a diferencias entre razas, especies y taxonomías más elevadas. La estructura de las poblaciones es tan compleja que una evaluación cuantitativa de su potencial evolutivo es extremadamente difícil. El crecimiento sin embargo es una característica susceptible de medición. El crecimiento ilimitado de la población es exponencial; sin embargo el ambiente tiende a estabilizarlo en un tamaño llamado "capacidad de acarreo" denotado por K . K es una variable que no solo depende de las capacidades reproductivas, sino también por el medio ambiente, la aptitud de sus depredadores y la abundancia de alimento entre otros. Estrategias tales como r -selección (incremento del número de descendientes) o K -selección (incremento en la ventaja selectiva de sus descendientes) permiten modelar las poblaciones observando los resultados aproximarse a la capacidad de acarreo.

El proceso de evolución natural es frecuentemente visto como una búsqueda sobre un espacio muy complejo, donde lo que se pretende encontrar es la configuración con los individuos más aptos. Esta visión original de Darwin de selección direccional ha sido cuestionada en años recientes, dando origen a otras hipótesis como la teoría de evolución neutra o preguntarnos sobre el origen del orden.

La teoría de evolución neutra [13] asegura que la mayoría de las substituciones están dadas por un "genetic drift" en lugar de por selección natural, es decir, por variaciones aleatorias en la codificación del individuo. Esta teoría no está en contradicción con la de Darwin, sino más bien asegura que la mayoría de las substituciones no tienen influencia en la supervivencia del genotipo, pero las pocas que sí, están sujetas a la adaptación de la especie al medio ambiente.

Una visión alternativa para la modelación de este fenómeno considera la evolución como un flujo por un terreno montañoso o "escenario de aptitud"; otra más considera a la evolución como una caminata adaptable sobre espacios con configuraciones de hipercubos N -dimensionales. Recientemente estas visiones han sido usadas para atacar el problema del origen del orden, dando lugar a un nuevo paradigma conocido como "generación espontanea de orden" [12].

La teoría de genética poblacional es mucho más madura que la de CE [4], por ello, algunos de los resultados obtenidos en la segunda son adaptaciones de la primera a un escenario artificial.

2.1.2. Evolución artificial

Tomando como base los extraordinarios resultados obtenidos por la naturaleza (seres extremadamente complejos como nosotros en solo millones de años), imitamos sus procesos con el fin de dar solución a problemas difíciles para métodos tradicionales. Iniciamos por imitar la representación de la información a través de cadenas o secuencias finitas de símbolos que hacen las veces de las cadenas de ADN. En el caso artificial generalmente hacemos uso del haploidismo, es decir una sola cadena contiene toda la información genotípica. Los alelos de estas cadenas están constituidos por símbolos de un alfabeto arbitrario, aunque es común emplear uno binario. La estructura básica de un algoritmo evolutivo se muestra en la figura 2.1 y consiste de los siguientes puntos:

Generación de la población inicial Se hace generalmente de manera aleatoria para garantizar que se cubra uniformemente el espacio de búsqueda, aunque se puede incorporar aquí el conocimiento previo del problema.

Evaluación de la aptitud Se aplica la función de aptitud a las cadenas para observar su capacidad y decidir así cuales son mejores.

Selección Se eligen las cuerdas que mejor se han desempeñado y estas son candidatos a reproducirse con mayor probabilidad.

Cruza Se eligen dos cuerdas y mediante un procedimiento preestablecido se generan otras 2 haciendo intercambios en los genotipos originales.

Mutación Se inducen pequeños cambios en los individuos para garantizar la exploración del espacio.

Verificación de la condición de fin La condición de paro puede ser el número máximo de ciclos o que la varianza de la aptitud sea menor a cierto valor entre otros criterios. De cumplirse terminamos el proceso y de manera contraria se reemplaza la antigua población con la nueva y se repite el proceso.

Una iteración de este ciclo es llamada una "generación". Entonces, la primera generación (generación 0) de este proceso es, como se dijo anteriormente, una población de individuos generados de manera aleatoria, aunque se pueden generar con base a conocimiento previo del problema o del espacio de búsqueda. A partir de este punto los operadores genéticos (selección, cruza y mutación) junto con la medida de aptitud, actúan para mejorar la población.

Como se mencionó en la introducción, hay tres principales divisiones en el estudio de los programas evolutivos: AGs, PG y EE. Este trabajo es sobre AGs canónicos así que abundaremos sobre ellos en la sección 2.5, sobre los otros dos a continuación damos una pequeña descripción:

PG es la extensión del modelo de aprendizaje genético al espacio de los programas. Esto es, los objetos que constituyen la población no son cadenas de longitud fija que codifican las posibles soluciones del problema; son programas



Figura 2.1: Diagrama de flujo de un AG

que, cuando son ejecutados “son” las soluciones candidatas al problema. Estos programas son expresados en PG como árboles de análisis o “parseo”, en lugar de líneas de código. Entonces, por ejemplo, el programa simple “ $a + b * c$ ” sería representado como en la figura 2.2, o para ser más precisos, como estructuras de datos adecuadas, ligadas para lograr este efecto. Dado que esto es muy fácil de conseguir en el lenguaje de programación Lisp, hay una tendencia por usarlo. Sin embargo, esto pertenece a los detalles de implementación. Hay métodos directos para implementar PG usando otros ambientes de programación. Los programas en la población están compuestos por elementos del “conjunto de funciones” (p. e. “+” y “*”) y el “conjunto de terminales” (ej. “a” y “b”), que típicamente son conjuntos fijos de símbolos seleccionados de manera apropiada para la solución de problemas en el dominio de interés.

Las EE son un tipo particular de algoritmo evolutivo desarrollado en Alemania en los 60’s. Emplea parámetros reales, y en su forma original se valía de la mutación como operador de búsqueda y tenía un tamaño de población 1. Desde entonces ha evolucionado a compartir muchas características con los AGs, aunque hay otras alternativas de representación como cadenas de longitud variable, listas ligadas, etc. Otro tipo de EE actualmente producen más descendientes que padres y después eliminan a los poco aptos.



Figura 2.2: Árbol de análisis para “ $a + b * c$ ”

2.2. Aptitud y escenarios de aptitud (espacio de estados/configuraciones)

En la naturaleza, la aptitud de un organismo es medida por el éxito del organismo en su vida. En el caso artificial la función de fitness (aptitud) determina el comportamiento del AG en el sentido de guiar la búsqueda. En general, sobre todo en sistemas naturales, esta función es dinámica (varía con el tiempo), induciendo cambios profundos en la configuración del escenario evolutivo que representan los cambios en el ambiente así como la evolución de otras especies. Los escenarios de aptitud son un modelo con base en la biología evolutiva, con los cuales podemos visualizar el escenario evolutivo como un terreno montañoso, y nuestro objetivo es siempre estar lo más alto posible, pensando que esto nos llevará eventualmente al “óptimo”. Cabe notar que en la naturaleza el concepto de óptimo no tiene sentido como tal, lo que tiene sentido es ser más apto que los demás de la población actual. En el caso artificial, la identificación entre el genotipo \mathcal{G} y el fenotipo \mathcal{F} es vía una función $\phi : \mathcal{G} \rightarrow \mathcal{F}$ dada según el problema particular. En el caso artificial la aptitud se mide con una función real no negativa $f : \mathcal{F} \rightarrow \mathbb{R}^+ \cup \{0\}$ que nos dice qué tan buena es una solución para el problema planteado. Dada una cadena $C \in \mathcal{G}$, por comodidad usaremos $f(C)$ en vez de $f(\phi(C))$ donde no haya lugar a confusión. En la figura 2.3 se muestran de manera esquemática estas transiciones.

En la literatura de genética poblacional hay acusaciones apuntado que la aptitud definida en términos Darwinianos es una tautología, o una definición circular. La razón es que desde esta óptica si nos preguntamos ¿quién es más apto? la respuesta es “el que sobrevive” y si preguntamos ¿quién sobrevive? la respuesta es “el más apto”. Una medida de aptitud usada con frecuencia en la evolución artificial es considerar simplemente la probabilidad de que un individuo sobreviva. Sin embargo, este tipo de medidas de aptitud dejan fuera de consideración varios factores relacionados con la forma en que se lleva acabo la reproducción. Una alternativa que sí considera estos factores es llamada “aptitud efectiva”.

Como mencionamos anteriormente, los cromosomas (cadenas de símbolos) de longitud N se pueden ver como puntos en un espacio de búsqueda N -dimensional

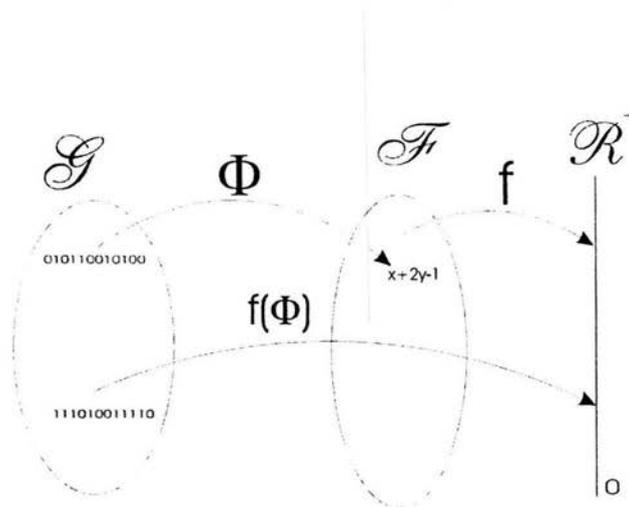


Figura 2.3: mapeo genotipo-fenotipo-aptitud

discreto, es decir, como vértices de un hipercubo. En la figura 2.4 vemos ejemplos de hipercubos para 1, 2 y 3 dimensiones. A continuación, a manera de ejemplo, mostramos la representación esquemática de los escenarios de aptitud que usaremos a lo largo de este documento. Junto a cada una de las cuerdas se muestra su aptitud entre paréntesis.

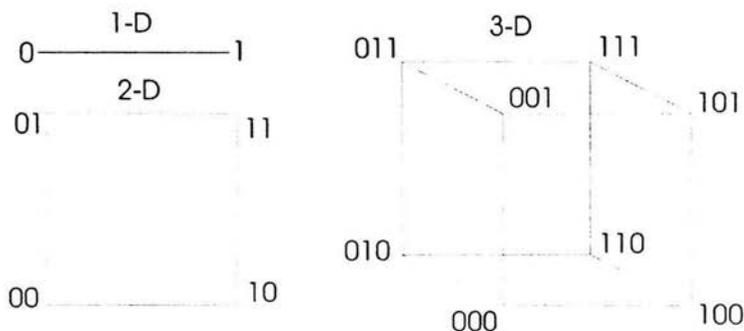


Figura 2.4: Interpretación geométrica de espacios de búsqueda de 1, 2 y 3 dimensiones

La noción de escenario de aptitud fue concebida originalmente por Sewall Wright [33] en los años 30. En el marco de la biología evolutiva, nos dan una visión intuitiva en la que la selección natural puede ser entendida en términos del flujo de una población en un terreno montañoso. En general los escenarios de aptitud son dinámicos por naturaleza, esta dinámica refleja los cambios que se presentan en el ambiente. Un ejemplo de esta variabilidad en el ambiente es la temperatura; si la temperatura aumenta en una región con clima frío, en-

tonces el más apto ya no será el que tiene más pelo para cubrirse, sino por lo contrario. Es decir, la cantidad de pelo pasa de ser una característica positiva a una potencialmente negativa. Considerar escenarios dinámicos en vez de estáticos, en general implica que el estudio se hará notablemente más complejo. Por esta razón inicialmente nos restringimos al estudio de escenarios estáticos y los usamos como primeras aproximaciones a los dinámicos.

Aún con la restricción de los escenarios al caso estático, el estudio de los AGs sobre escenarios de aptitud generales es muy complicado. Por esto restringimos aún más el problema considerando como etapa inicial algunas idealizaciones y casos límite representativos de las diferentes clases de problemas. Colin Reeves presenta una caracterización matemática del término escenario de aptitud, iniciando desde la definición de vecindad, lo que nos permite tener un panorama más amplio del tema [19]. Stadler y Stephens analizan la utilidad de los escenarios de aptitud como herramienta para analizar la dinámica de los procesos evolutivos [22].

El caso más sencillo es el escenario plano. En éste $f(C_i) = r, \forall C_i$ cadena en la población y con r constante real positiva. Podemos visualizar este escenario como el límite cuando la presión selectiva es muy débil, así podemos estudiar modelos de evolución neutral. En la práctica nos encontramos en este escenario por ejemplo cuando el algoritmo está cerca a la convergencia y las cuerdas en la población cada vez tienen una aptitud más homogénea. Aunque pudiera parecer una sobresimplificación, el escenario plano nos da un punto de partida para el análisis de escenarios más generales. Cabe notar que muchas veces es el único caso en el que se encuentran soluciones exactas. En la figura 2.5 vemos el cubo para el escenario plano. Notemos que no tiene mucho sentido hablar de la existencia de óptimos. Aquí la evolución no es guiada por selección, esto ocurre por ejemplo cuando las diferencias en las aptitudes de los individuos son pequeñas (no importando que los individuos sean muy diferentes). Vose [30] entre otros ha estudiado las regiones en las que no hay presión selectiva, que son conocidas como redes neutrales.

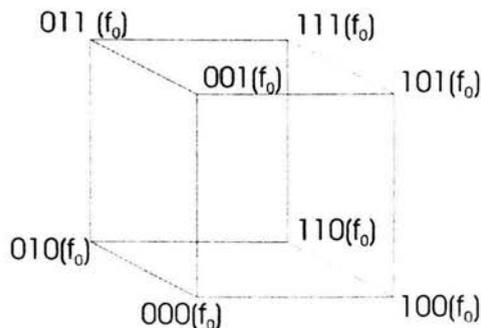


Figura 2.5: Cubo para un escenario plano

El siguiente escenario que consideramos es el escenario de contando unos, como su nombre lo indica definimos la aptitud de una cadena como el peso de Hamming de la misma ($f(C_i) = w_H(C_i)$). Este escenario encapsula una gran cantidad de problemas en los que la función de aptitud se comporta de manera quasi-lineal, es decir podemos optimizar cada variable independientemente. Jon Rowe ha hecho un estudio de este tipo de funciones en [21]. En general tenemos $f(C_i) = \sum_{k=0}^N f_b(C_i^k)$, donde $f_b(C_i^k)$ denota la aptitud del k -ésimo bit. A este tipo de escenarios podemos verlos también como el límite cuando tenemos cero epistasis, la epistasis es el término para las relaciones no lineales que existen entre las variables. En la figura (2.6) observamos la existencia de un máximo global (111) y un mínimo global (000).

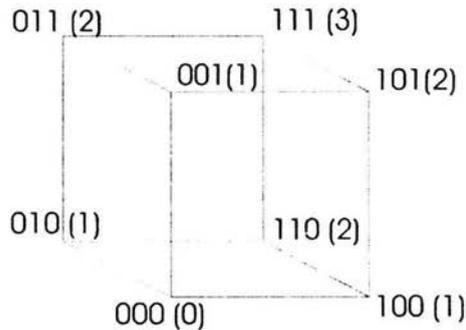


Figura 2.6: Cubo para un escenario de contando 1's

Otro escenario de cierta importancia teórica es el que nos brinda el modelo de Eigen o "needle-in the-haystack" [6]. El modelo de Eigen consiste en considerar un escenario en el que todos los genotipos tienen la misma aptitud ($f(C_i) = f_0 \forall C_i \neq C_{sm}$) salvo uno, que resulta ser el óptimo ($f(C_{sm}) = f_{sm}$), a este punto lo llamamos la secuencia maestra. Este modelo nos permite apreciar el caso en el que los puntos del escenario están altamente no correlacionados. Es decir, el escenario no nos brinda información relevante acerca de la posible localización del óptimo. En este caso no podemos hablar de pendiente selectiva, es decir, es el caso límite cuando la población no nos permite hacer explotación de las soluciones candidatas actuales (salvo en presencia de la secuencia maestra). En la figura (2.7) apreciamos la existencia de un máximo (la secuencia maestra 001) con aptitud f_{sm} , pero todos los demás puntos del escenario comparten el mismo valor en la función de aptitud f_0 , así que la búsqueda se convierte en un proceso de encontrar la aguja en el pajar (needle-in-a-haystack), o en otros términos, un proceso en el que solo hay exploración hasta que se encuentra la solución.

Otro escenario que consideraremos, aunque con menos profundidad es el escenario de contando Kinks que se define mediante la función

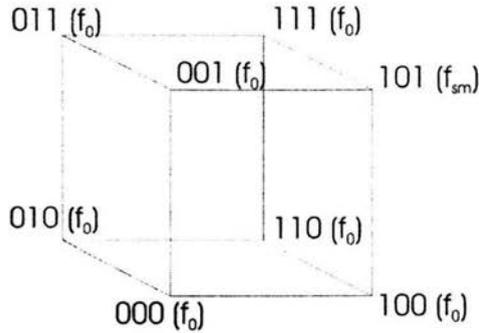


Figura 2.7: Cubo para un escenario del modelo de Eigen con secuencia maestra 001

$$f(C_i) = \sum_{j=0}^{N-1} \sigma_j \sigma_{j+1} \quad (2.1)$$

Donde σ_j es -1 si el j -ésimo bit de la cadena es 0 y tiene el valor de 1 si ese mismo bit es 1. En otras palabras contamos el número de cambios entre los bits consecutivos. Realmente llamamos un kink al cambio de un bit con valor 1 a uno con valor 0 y un antikink a lo opuesto, pero en este caso solo consideramos los cambios sin importar la “dirección”. Este escenario tiene importancia pues presenta epistasis, aunque su estudio es complicado. El ejemplo para el caso de 3 bits se muestra en la figura 2.8, como podemos observar este escenario cuenta con 2 mínimos (101,010) y 2 máximos (000,111)

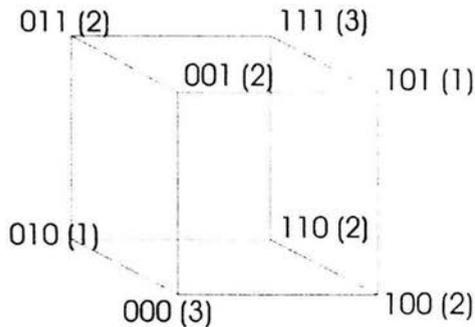


Figura 2.8: Cubo para un escenario de contando Kinks

2.3. Dinámica

La dinámica de un sistema está dada por un cambio de estado del tiempo t al $t + 1$. Este cambio está dado por un sistema de reglas de producción definidas para un espacio de configuraciones, pero las reglas de producción deben ser aplicadas a un espacio de búsqueda particular para su estudio. Es decir, solo cuando combinamos las reglas con un espacio de búsqueda es que obtenemos un comportamiento dinámico. Por lo tanto para observar el comportamiento de un AG necesitamos probar sus reglas de producción (operadores genéticos) sobre diferentes espacios de búsqueda para poder aprender sobre la dinámica. Podemos decir a grandes rasgos que: escenarios de aptitud + operadores genéticos = dinámica de los AGs.

En el caso especial de AGs, un estado del espacio de configuraciones está definido por el subespacio particular que forma la población actual (equivalentemente por las proporciones de cada cadena). Y las reglas de transición están dadas por los operadores genéticos. Estos últimos son los que generan un movimiento de dicha población.

Un ejemplo de un operador que genera dinámica sobre escenarios de aptitud está en la propuesta de Kauffman [12]; en esta propuesta la dinámica está dada por caminatas aleatorias sobre hipercubos N -dimensionales. Las caminatas consisten en inducir pequeñas mutaciones a un individuo (que en este caso es una población), esperando que esto nos lleve eventualmente a la solución óptima. La evolución aquí puede ser visualizada como un camino entre los vértices del hipercubo que nos lleva a mejores soluciones (Hill Climbing). Con esta óptica podemos apreciar la preferencia que tienen las mutaciones positivas. Desafortunadamente en la práctica este modelo tiende a estancarse en óptimos locales con gran facilidad. Podemos encontrar una revisión y extensión al trabajo de Kauffman en el trabajo de Lee Altenberg [2].

En la siguiente tabla mostramos un ejemplo del modelo Nk de Kauffman con $N = 3$ y $k = 2$. Generamos aleatoriamente el valor para cada uno de los 3 bits w_1 , w_2 y w_3 , con esto se pretende crear el efecto de relaciones no-lineales desconocidas entre los bits. Se define entonces la aptitud como el promedio de estos valores, esto genera un escenario de aptitud con múltiples óptimos locales. En la figura 2.9 podemos observar gráficamente el escenario de aptitud de la tabla 2.1 donde 111 y 100 están marcados por un círculo que denota un óptimo local.

A continuación mostramos nuevamente los escenarios de aptitud típicos con los que trabajaremos, esta vez introduciendo el efecto de la dinámica. La dinámica está dada por la aparición de mutaciones de 1 bit por generación, así podemos ver la trayectoria desde cualquier punto hacia el óptimo. En los diagramas se muestran flechas que indican la dirección del flujo de la población, dado por la presión selectiva.

En el caso de un escenario plano (figura 2.10) vemos un cubo en el que apreciamos que no hay ninguna posibilidad de un flujo definido hacia algún punto en particular. En lugar de esto, hay una distribución aleatoria de los individuos en todo el espacio, a diferencia del escenario de contando 1's (fig

Cuadro 2.1: Valores de la aptitud de cada gene

1	2	3	w_1	w_2	w_3	$W = \frac{1}{N} \sum_i^N w_i$
0	0	0	0.6	0.3	0.5	0.47
0	0	1	0.1	0.5	0.9	0.50
0	1	0	0.4	0.8	0.1	0.43
0	1	1	0.3	0.5	0.8	0.53
1	0	0	0.9	0.9	0.7	0.83
1	0	1	0.7	0.2	0.3	0.40
1	1	0	0.6	0.7	0.6	0.63
1	1	1	0.7	0.9	0.5	0.70

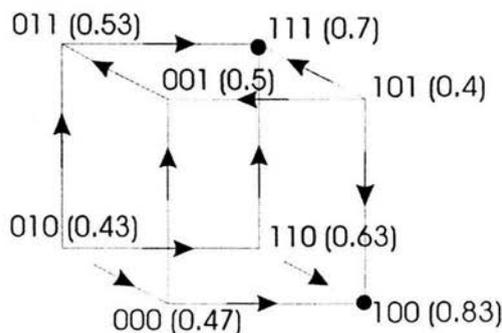


Figura 2.9: Dinámica de un escenario del modelo de Kauffman

2.11) en la que apreciamos una ruta bien definida desde cualquier punto hacia el óptimo. El siguiente caso es el modelo de Eigen (fig. 2.12), en el que no hay forma de saber el rumbo hacia el óptimo salvo al estar a distancia de Hamming 1. Escenarios más interesantes son los de contando kinks (fig. 2.13) y el de Kauffman (fig. 2.9)

Siempre podemos calificar un algoritmo de búsqueda en espacios complejos mediante el uso de 2 conceptos básicos que son la exploración y la explotación. La exploración se refiere a probar nuevas rutas en la búsqueda, mientras que la explotación se refiere a qué tan lejos llegamos a profundizar dichas rutas esperando encontrar el óptimo. En analogía con los heurísticas de búsqueda en árboles de decisión, la exploración corresponde a búsqueda en amplitud mientras que la explotación corresponde a una búsqueda en profundidad. Un algoritmo heurístico de búsqueda ideal debe tener una combinación equilibrada de estos dos elementos de tal manera que nos permita tener la certeza de que se explorará el espacio con cierta regularidad y que se han explotado los subespacios más prometedores para así encontrar el óptimo global. Notemos que para encontrar la solución óptima a un problema es suficiente la exploración; sin embargo puede tomar mucho tiempo examinar cada solución. Necesitamos entonces del meca-

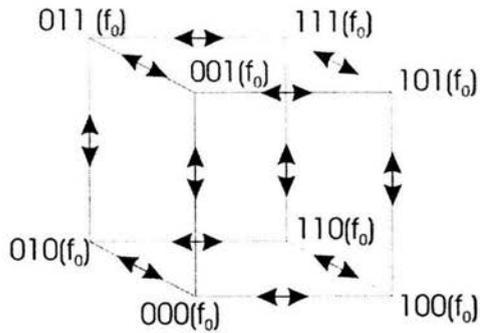


Figura 2.10: Dinámica de un escenario plano

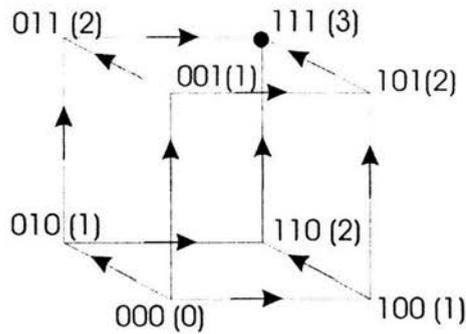


Figura 2.11: Dinámica de un escenario de contando 1's

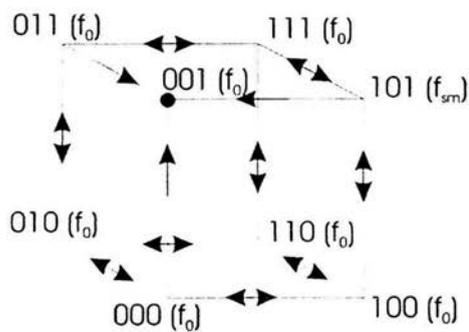


Figura 2.12: Dinámica de un escenario del modelo de Eigen con secuencia maestra 001

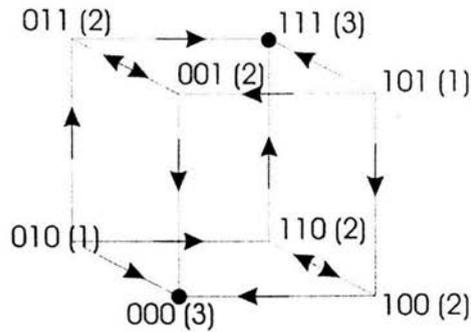


Figura 2.13: Dinámica de un escenario de contando Kinks

nismo de explotación para acelerar la búsqueda considerando la información del escenario en particular.

Un ejemplo típico en donde podemos examinar el antagonismo entre exploración y explotación es el del tragamonedas [11]: supongamos que tenemos una máquina tragamonedas de 2 brazos de la cual queremos obtener el mayor provecho, supongamos también que de alguna manera sabemos que la ganancia en un brazo es mayor (en promedio) a la del otro. Nuestro objetivo es: por un lado investigar con cierta precisión cual es el brazo que nos dará más ganancia, minimizando por supuesto el número de intentos en el brazo equivocado, y por el otro lado queremos maximizar el número de intentos en el brazo correcto. Dado que tenemos un número finito de intentos, tenemos el problema de qué tanto vamos a intentar en cada brazo antes de decidir cual es el deseado (exploración) y qué tanto vamos a usar el brazo elegido hasta el momento (explotación). Si dedicamos demasiado tiempo a averiguar con certeza cual es el brazo adecuado, no tendremos oportunidad de usarlo. Pero si prematuramente explotamos uno de ellos creyendo que es el indicado, tendremos incertidumbre de si era realmente el indicado.

En general es muy difícil conseguir este tipo de heurísticas dada la gran variabilidad de los espacios de búsqueda, dado que los espacios de interés son complejos, es decir están en el umbral entre lo simple y lo caótico. Un error común es el comparar a los AGs con otros métodos de búsqueda sobre espacios estáticos, pues los mayores beneficios se observan al considerar espacios dinámicos en el tiempo. Los AGs ofrecen una excelente combinación entre exploración y explotación, se ha demostrado con la práctica que son muy buenos para encontrar la región en la que se encuentra el óptimo global. Sin embargo, no localizan el óptimo con la misma velocidad.

2.4. Cambio de escala

El concepto de “coarse graining” o cambio de escala, presentado en este contexto, nos da la sensación de ver el problema con una perspectiva más lejana, en muchas ocasiones no necesitamos saber todos los detalles del problema para darle solución, incluso, demasiados detalles pueden traer complicaciones innecesarias. El observar el problema desde diferentes grados de detalle nos permite encontrar los grados efectivos de libertad del problema en particular. Un caso particular de cambio de escala son los esquemas que han sido empleados de manera generalizada para modelar los AGs. Dada su importancia hablaremos de ellos más adelante.

El concepto de cambio de escala es empleado ampliamente en distintas áreas. En la física por ejemplo, la temperatura es el promedio de la energía cinética de las partículas; sin embargo resulta mucho más útil hablar de la temperatura como un solo valor que trabajar con una cantidad inmensa de energías cinéticas para cada partícula. Algo similar ocurre con la presión, es por eso que simplemente usamos estas medidas en lugar de los 10^{23} grados de libertad de las partículas. En economía tomamos como indicador el índice de precios y cotizaciones (IPC) en lugar de preocuparnos por cada una de las transacciones por separado; es decir, el IPC es una variable trascendente después del cambio de escala. Para el estudio de la economía de un país por ejemplo, no es práctico hablar del movimiento de capital de cada persona, en su lugar usamos algunos grados de libertad efectivos como “el sector turismo” o el “sector automotriz”. Entonces podemos hablar por ejemplo de que hay una dinámica efectiva sobre las variables “turismo” y “automotriz” (después del cambio de escala).

Hay sistemas en los que los grados de libertad son obvios. Por ejemplo, en economía podemos crear toda la taxonomía³ iniciando desde la economía mundial, seguida por la división en la economía de los países que a su vez está dividida en sectores automotriz, turismo, etc. y así podemos continuar hasta hacer una división por personas. Sin embargo los sistemas complejos tienen por característica que es difícil encontrar los grados de libertad efectivos. Por ejemplo, considerando el problema del agente viajero, quizá hay una base adecuada en la que su resolución es trivial, si conociéramos dicha descomposición probaríamos $P=NP$. Esto nos da una idea de la complejidad que involucra identificar los grados de libertad.

Para estudiar organismos naturales requerimos de mecanismos de clasificación que nos permitan reconocer sus características relevantes. Así creamos un árbol con dichas clases y podemos estudiar a diferentes profundidades los organismos. Por ejemplo podemos estudiar el comportamiento de una persona en particular, podemos estudiar la raza a la que pertenece, podemos estudiar la especie ser humano, etc. En la taxonomía biológica, la raíz del árbol de clasificación es “ser vivo” y los nodos hoja son cada uno de los individuos, pero hay muchos niveles intermedios. Tenemos la distinción entre reinos y otras clases intermedias hasta llegar a especies, razas, género y por último individuos

³jerarquización de los grados de libertad en diferentes escalas

particulares.

Una de las motivaciones más importantes para el uso de variables escaladas en biología es la gran cantidad de grados de libertad que aparecen. Por ejemplo, una proteína está constituida por una cantidad del orden de 10^4 aminoácidos por lo que tenemos aproximadamente 22^{10^4} configuraciones. Por otro lado el universo tiene aproximadamente 10^{80} átomos y $22^{10^4} \gg 10^{80}$; esto quiere decir que el tiempo que tomaría examinar todas estas configuraciones⁴ excedería por mucho la edad calculada del universo actual. En el caso artificial, encontramos otro ejemplo en el que la cantidad de configuraciones crece de manera inmanejable, es el problema del agente viajero, con solo 100 ciudades el número de configuraciones es $100!$. Como podemos observar, aún problemas que podemos llamar “pequeños” tienen demasiados estados para considerarlos de manera separada. Este es otro motivo para considerar variables escaladas.

Al estudiar evolución, dado que no tiene sentido preguntarnos como evoluciona cada individuo por separado, consideramos poblaciones de diferentes tamaños, por lo general agrupamos todos los individuos de la misma especie y los observamos como un todo. Es por esto que la evolución es un estudio sobre las poblaciones nunca sobre individuos⁵, una manera de agrupar individuos es mediante el uso de un criterio de similitud, en el caso artificial definimos una relación (en el sentido matemático) para agrupar cadenas con características comunes (que tengan cierto allele en cierto locus o que tengan el mismo número de unos). Esta agrupación nos permite saber como se comporta cierto grupo de genes y por qué algunos tienen mayor repercusión en la función de aptitud.

2.5. Algoritmos genéticos

Los AGs son métodos estocásticos de búsqueda basados en poblaciones que haces uso de la metáfora natural de la evolución. Básicamente es un proceso iterativo en el que consideramos un conjunto de soluciones candidatas a un problema dado, al que llamamos una población. Estas soluciones (individuos o cromosomas) compiten bajo ciertas reglas (operadores genéticos) para evolucionar y preservar sus características (genotipo). Si tienen características deseables (con aptitud mayor al promedio) tienen la oportunidad de reproducirse y su descendencia pasa a la siguiente fase (generación). Este proceso se repite hasta que cierto criterio se satisface. Por ejemplo si se alcanza un número máximo de iteraciones (generaciones) o la función de aptitud de algún individuo sobrepase una cierta cota preestablecida.

En AGs consideramos poblaciones de tamaño fijo a diferencia de las EE en las que se permite un tamaño variable temporalmente. Esto podría parecer una desventaja, sin embargo es una buena aproximación al fenómeno natural en el que una especie en particular llega a un equilibrio en su población causado por limitaciones de recursos y muertes por depredadores. Una vez alcanzado

⁴con una computadora de un Gigaflap

⁵Esto a pesar de que la selección natural actúa fundamentalmente a nivel de los individuos

el equilibrio hay un intervalo de tiempo en que el número de integrantes de la población casi no cambia.

Un campo en el que se ha difundido la utilización de AGs es el de los sistemas complejos. Los sistemas complejos se encuentran en el umbral entre lo simple y lo caótico; en este umbral se encuentran los problemas que podemos considerar “interesantes”. Algunas características distintivas que presentan estos sistemas son:

- Un número grande de grados de libertad. Los grados de libertad son las variables a considerar para describir un fenómeno. Como su número en general es muy grande, es importante identificar los grados de libertad efectivos. Es decir las variables relevantes que nos llevarán a resolver el problema.
- Presentan epistasis. Es un término tomado de la literatura biológica y se refiere a interacciones no lineales entre las variables o grados de libertad. Las interacciones entre las variables de un problema en general son desconocidas. Las interacciones no lineales entre las variables dificultan encontrar el óptimo de manera progresiva. Es decir, la epistasis no nos permite optimizar variables una a una.
- Frustración. Se refiere a cuando algunas restricciones en el problema resultan contradictorias. En los problemas que presentan frustración la información local del escenario (en forma de gradiente) es engañosa y por lo tanto puede retrasar o incluso evitar la convergencia al óptimo.
- Jerarquización. Existe una estructura de árbol en los grados de libertad efectivos y el problema presenta muchos niveles a los cuales puede ser tratado.

Los AGs son frecuentemente vistos como optimizadores de funciones, pero su uso es mucho más extendido, por ejemplo como máquinas de búsqueda, clasificadores, sistemas de aprendizaje entre otros.

2.5.1. Cuerdas

Las cuerdas son los elementos primordiales de los AGs canónicos, estas son la representación de las soluciones al problema en cuestión, es decir que la misma cuerda tiene diferentes interpretaciones (los parámetros para una red neuronal o el orden en el que se visitan las ciudades en el problema del agente viajero), que son asociadas a la función de aptitud. A continuación se presenta una serie de definiciones formales para el estudio de las cuerdas:

Consideremos un conjunto finito de símbolos Ω al que llamaremos alfabeto. Definimos el conjunto de cuerdas de longitud N sobre Ω como: Ω^N . Con base en esto, definimos una población como un multiconjunto (multiset) de cuerdas tomadas de Ω^N , para fines prácticos limitamos la cardinalidad de la población (denotada por n), manteniéndola constante, al igual que a N a lo largo del

algoritmo. Aunque esto no es una restricción para su funcionamiento, nos ayuda a simplificar el estudio y es lo común en la práctica. Notemos que el número de estados posibles para una población es $n|\Omega|^N$, donde $||$ denota la cardinalidad.

Dada una población a un tiempo t nos interesa medir el número de apariciones que tiene una cuerda C_i , denotaremos este número por $n(C_i, t)$. Este término será usado igualmente para denotar el número esperado de cadenas C_i al tiempo t . En consecuencia, definimos la proporción esperada de la cuerda C_i al tiempo t en la población como $P(C_i, t) = n(C_i, t)/n$, considerando a $P(C_i, t)$ como el elemento fundamental de nuestro estudio.

Debemos resaltar que en el límite cuando el tamaño de la población tiende a infinito, $P(C_i, t)$ es la probabilidad de que la cuerda C_i exista en la población, al tiempo t . En el caso de una población finita tendremos algunas variaciones debidas a los procesos estocásticos como la selección proporcional.

2.5.2. Representación

Denotamos por $\mathbf{P}(t)$ al vector de proporciones $\mathbf{P}(t) = (P(C_0, t), \dots, P(C_{|\Omega|^N-1}, t))$. Si el alfabeto es binario la dimensión de $\mathbf{P}(t)$ es 2^N . La dinámica del sistema es inducida por una regla de transición τ que mapea la población de una generación a otra $\mathbf{P}(t) \xrightarrow{\tau} \mathbf{P}(t+1)$. No todas las reglas de transición son válidas. Una buena representación para la población es un primer paso hacia la caracterización de los τ admisibles. Definamos el simplex como el conjunto

$$\Lambda = \{(P(C_0, t), \dots, P(C_{|\Omega|^N-1}, t)) : \sum_{i=0}^{|\Omega|^N-1} P(C_i, t) = 1, P(C_i, t) \geq 0\}$$

Para tener una idea geométrica del espacio de representación, presentamos Λ en la figura 2.14 para los casos particulares de 2 y 3 dimensiones.

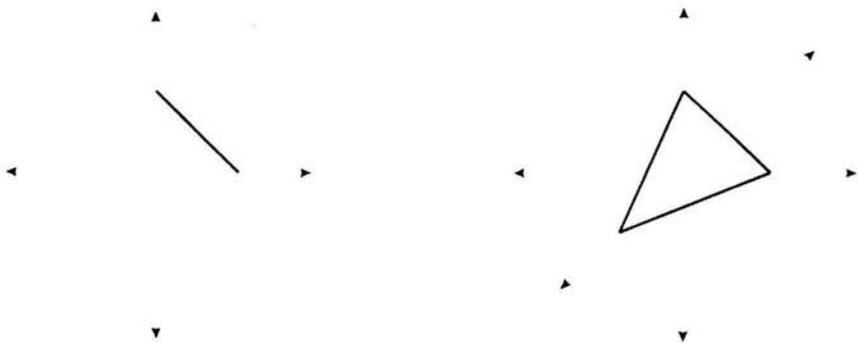


Figura 2.14: Simplex de dimensión 2 y 3

Notemos que cada vértice de Λ corresponde a un vector unitario básico del espacio. Por ejemplo, los vértices de la figura de la derecha son los vectores

básicos $(0,0,1)$, $(0,1,0)$ y $(1,0,0)$ que corresponden a una población homogénea y $(1,0,0)$ representa que solo hay cuerdas de tipo C_1 en ese momento. En el caso de AGs canónicos, Λ es un tetraedro de dimensión $2^N - 1$ contenido en un espacio de dimensión 2^N .

2.5.3. Operadores genéticos

Llamaremos operador genético a aquella función que mapea una población en otra. En la literatura de AGs una gran cantidad de operadores genéticos se han propuesto. En el presente documento nos concentraremos en los tres clásicos que también son los más usados; estos son: selección, mutación y cruza.

El operador de selección consiste en obtener los individuos más aptos (con función de aptitud más alta) de una población dada. La cruza consiste en hacer una mezcla de los genotipos de individuos produciendo nuevos (descendencia) esperando que las características heredadas de cada padre en suma tengan un mejor valor en la función de aptitud. Otro operador común es el de mutación, que consiste en inducir pequeñas variaciones en el genotipo del individuo con el fin de diversificar la información. En general un operador genético es de la forma:

$$\mathbf{P}(t+1) = \mathcal{H}\mathbf{P}(t) \quad (2.2)$$

donde \mathcal{H} es un operador (en el sentido matemático) en el espacio de búsqueda. Este operador en un contexto general es no lineal y depende de: las proporciones de todas las cadenas, de los valores de aptitud de las cadenas y de las probabilidades de cruza y mutación entre otras. Es decir:

$$\mathcal{H} = \mathcal{H}(\{C_i\}_i, \{p_k\}_k, \{f_i\}_i, t), i \in [0, |\Omega|^N - 1] \quad (2.3)$$

Con $\{C_i\}_i$ el conjunto de cadenas, $\{f_i\}_i$ el conjunto de evaluaciones sobre dichas cadenas de la función de aptitud y $\{p_k\}_k$ un conjunto de parámetros.

Selección

Selección se refiere, en términos Darwinianos a la sobrevivencia de los más aptos. En el caso de la evolución artificial, declaramos para eso una función de aptitud, en base a la cual podemos decidir que individuos se desempeñan mejor y en base a eso podemos decidir si pasa a la siguiente generación o no.

Un método de selección muy común es el conocido como “selección por rueda de ruleta” o selección proporcional. Podemos visualizar este método como sigue: supongamos que tenemos una superficie circular, la dividimos en secciones de tal manera que el área de la sección i es proporcional a la aptitud de la cuerda i . Ahora “ponemos a girar” la ruleta, y dependiendo de la sección en la que deje de girar habremos seleccionado una cuerda que pasará a la siguiente generación. Repetimos esta operación tantas veces como elementos tengamos en la población. Claramente esta forma de selección estocástica hace que las probabilidades de seleccionar cuerdas aptas sea alta. Notemos que dada la naturaleza aleatoria de esta forma de selección, podemos perder incluso a la más apta (aunque esto

puede ser poco probable). Esta analogía es equivalente a decir formalmente que el valor de aptitud de una cuerda dividido entre la suma de la aptitud de todas las cuerdas de la población actual es la probabilidad de que la cuerda pase a la siguiente generación.

Un σ -Torneo es otra forma de seleccionar individuos de una población que consiste en elegir σ individuos de la población de manera aleatoria y de ellos el que tiene mayor función de aptitud es el que pasa a la siguiente generación. Repetimos el proceso tantas veces como elementos hay en la población.

Hay dos maneras en que una propiedad \mathcal{P} aparezca en la población. La primera es que bajo el efecto de adaptación, es decir, que la población no se mantenga acumulada alrededor de un pico, sino que se disperse en un volumen grande del escenario de aptitud. Entonces la probabilidad que los individuos tengan la propiedad \mathcal{P} es alta y la segunda es que aunque la población se quede dentro de un volumen pequeño del escenario, \mathcal{P} es típica del escenario entero.

Denotaremos a la proporción de cuerdas de tipo C_i a tiempo t después de la selección por $P'(C_i, t)$.

Mutación

Este operador consiste en inducir “pequeñas” variaciones a los individuos de la población con cierta probabilidad. Para cada cadena C_i el gene C_i^k será alterado con probabilidad $p_m(k)$, dichas alteraciones dependerán del alfabeto. Por ejemplo cuando consideramos un alfabeto binario, esta alteración consiste en una sustitución por el otro elemento de dicho alfabeto. Es importante tomar en cuenta los casos especiales para p_m que se muestran en la tabla 2.2.

Cuadro 2.2: Casos de interés para los valores de p_m

p_m	Efectos
pequeño (~ 0)	Poca variabilidad permite la explotación de soluciones prometedoras pero hace lenta la convergencia al óptimo global y hay mayor probabilidad de estancamiento en óptimos locales
$\sim \frac{1}{2}$	Demasiados cambios no permiten conservar características deseables. No permite la explotación de las soluciones buenas obtenidas hasta el momento. Es el límite cuando se tiene pura exploración
~ 1	Tiende a invertir las cuerdas, lo que permite la exploración de soluciones muy diferentes a las actuales. Sin embargo invertir una cuerda apta puede resultar en una mala solución y podríamos perder la solución original

Debemos notar que, además de dar variedad a las nuevas generaciones a través de la creación de cromosomas inexplorados, el operador de mutación sirve como una especie de seguro. Es decir, nos permite recuperar bits perdidos con anterioridad y que no podríamos recuperar vía selección ni cruza. Una causa

común de pérdida de información es conocida como efecto de tamaño finito y se refiere a los errores en la aproximación que hicimos sobre una población infinita. El operador de mutación es completamente de exploración. En las ecuaciones denotaremos por $P_m(C_i, t)$ a la proporción de cuerdas de tipo C_i a tiempo t después de la mutación.

Cruza

El operador de cruce se utiliza para combinar características de individuos aptos esperando obtener de ellos las mejores características y transferirlas a la siguiente generación mediante su descendencia. De esta manera se obtiene diversidad dentro de la nueva población y a la vez se preservan las características deseables de los individuos.

Este operador normalmente es binario (es decir, toma dos argumentos) así que iniciamos el proceso de cruce seleccionando dos cadenas $C_i = c_i^1 c_i^2 \dots c_i^N$ y $C_j = c_j^1 c_j^2 \dots c_j^N$; estas cadenas se cruzarán con probabilidad p_c y su descendencia pasará a la siguiente generación. De otra manera las cuerdas pasan a la siguiente generación sin ser alteradas.

Por ejemplo, en el caso de la llamada cruce de un punto el procedimiento es el siguiente: una vez que tenemos las cadenas que se cruzarán, elegimos de manera aleatoria un número k , $0 < k < N$ al que llamaremos punto de cruce; entonces a partir de C_i y C_j generamos las cuerdas C'_i y C'_j (a las que llamaremos la descendencia de C_i y C_j) de la siguiente manera: $C'_i = c_i^1 c_i^2 \dots c_i^{k-1} c_j^k \dots c_j^N$ y $C'_j = c_j^1 c_j^2 \dots c_j^{k-1} c_i^k \dots c_i^N$. Por ejemplo si $C_i = 11001011$, $C_j = 11011111$ y $k = 5$, entonces $C'_i = 11001111$ y $C'_j = 11011011$. De manera natural extendemos este concepto a cruce de r puntos de la siguiente manera: elegimos $0 < k_1 < \dots < k_r < N$; entonces la cadena C'_i estará compuesta por bloques de las cadenas C_i y C_j alternadamente: $C'_i = C_i^1 \dots C_i^{k_1} C_j^{k_1+1} \dots C_j^{k_2} C_i^{k_2+1} \dots C_i^{k_3} C_j^{k_3+1} \dots$ y el segundo descendiente $C'_j = C_j^1 \dots C_j^{k_1} C_i^{k_1+1} \dots C_i^{k_2} C_j^{k_2+1} \dots C_j^{k_3} C_i^{k_3+1} \dots$. Un ejemplo para el caso de 2 puntos es: si $C_i = 11001011$, $C_j = 01011111$, $k_1 = 2$ y $k_2 = 5$, entonces $C'_i = 11011011$ y $C'_j = 01001111$. Por último consideramos la llamada cruce uniforme que consiste en elegir el gene k de la cadena C_i con probabilidad $p_u(k)$ y de la cadena C_j con probabilidad $1 - p_u(k)$. Por ejemplo si $C_i = 01011011$ y $C_j = 01001011$, entonces unos posibles valores para C'_i y C'_j son: $C'_i = 01001011$ y $C'_j = 01001011$

Podemos extender estos resultados a un esquema más general de cruce en el que consideramos máscaras de cruce. Una máscara es una manera de tomar bits de un padre, de modo que el complemento sean los bits que deberemos tomar del otro. En términos formales, una máscara es una partición en 2 del conjunto de índices de los bits de una cadena. Cada elemento de la partición corresponde a los bits que aportará el padre respectivo a la descendencia. Denotaremos por s y c a los elementos de esta partición, entonces $s = \{i_1, i_2, \dots, i_{N_s}\}$ y $c = \{j_1, j_2, \dots, j_{N_c}\}$. s y c deben cumplir $s \cup c = [0, N - 1]$ y $s \cap c = \{\}$, por lo tanto $N = N_s + N_c$ y $j_l \neq i_k \forall l, k$. Dada esta definición es claro que la cruce de r puntos es un caso particular de máscaras.

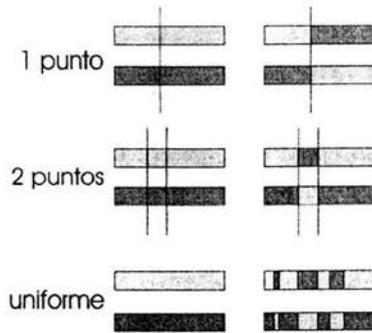


Figura 2.15: Tipos comunes de cruza, a la izquierda los padres y a la derecha su descendencia dependiendo del tipo de cruza. Las líneas verticales indican los puntos en los que se hace el intercambio.

El operador de cruza tiene, como veremos más adelante, el efecto de acelerar la convergencia del algoritmo pues permite tanto explorar como explotar soluciones. Denotaremos por $P_c(C_i, t)$ a la proporción de cuerdas de tipo C_i a tiempo t después de aplicar el operador de cruza.

2.6. Antecedentes sobre la evolución artificial

Con el fin de formalizar el estudio de los AGs se han propuesto varios modelos; el primero y más ampliamente difundido por su naturalidad es el de esquemas [11]. Los esquemas son una formalización en la que se añade un símbolo (representado generalmente como “*”) al alfabeto. El símbolo “*” es llamado “don’t care” o “no importa”. Si encontramos este símbolo en una cadena del nuevo alfabeto (a la que llamaremos un esquema) querrá decir que esa entrada debe ser ignorada; esto nos permite concentrarnos en las partes de las cadenas que nos interesan. Es decir, los esquemas nos permiten hacer una división en subproblemas. Esta visión está muy ligada con la hipótesis conocida como *hipótesis de los bloques constructores* propuesta por Goldberg [10]. La hipótesis de bloques constructores dice que los esquemas compactos y con aptitud mayor al promedio tienden a perdurar en el tiempo y unirse para construir soluciones con aún mayor aptitud que eventualmente convergerán a la solución.

Un resultado teórico importante que durante muchos años se ha considerado como la prueba de que los AGs funcionan es el famoso Teorema de esquemas (Holland 1975 [11], Goldberg 1989 [10]). El Teorema de esquemas básicamente dice que los esquemas con valor de aptitud mayor al promedio reciben un número “exponencial” de pruebas a través del tiempo y está dado en forma de la siguiente desigualdad.

$$P(\xi, t+1) \geq P'(\xi, t) \left(1 - p_c \left(\frac{l(\xi)}{N-1} \right) - N_2 p_m \right) \quad (2.4)$$

con $l(\xi)$ la longitud definitoria de un esquema de la cual hablaremos en la sección 4.3. El teorema de esquemas en su forma tradicional (2.4) solo nos da una cota inferior para la proporción esperada del esquema ξ . La razón es que no toma en cuenta el efecto de reconstrucción de esquemas; este efecto resulta ser de importancia como se muestra en [28].

Otra forma de modelar el comportamiento de los AG está dada por la propuesta de Vose [30]. Como se muestra en [31], el modelo del AG Simple es introducido por Vose y Liepins en 1991 [30]. En éste modelo, el vector $s^t \in \mathbb{R}^N$ representa la t -ésima generación del AG y el i -ésimo componente de s^t es la probabilidad de que la cadena i sea seleccionada del "pool genético". En esta formulación,

$$s_i^t \sim P(C_i, t) f(C_i) \quad (2.5)$$

donde \sim es la relación de equivalencia tal que $x \sim y$ si y solo si $\exists \gamma > 0$ tal que $x = \gamma y$. El término $1/\bar{f}$, que representa el promedio de la función de aptitud normalmente asociada con la reproducción proporcional, puede ser absorbido por el término γ , en general $\gamma = \gamma(t)$.

El número esperado de cadenas también es representado como el vector $p^t \in \mathbb{R}^N$ donde el k -ésimo componente del vector es igual a la representación proporcional de k en la generación t antes de que la selección ocurra. Esto sería lo mismo que $P(C_k, t)$ en una notación más comúnmente asociada al teorema de esquemas. Finalmente sea $r_{ij}(k)$ la probabilidad de que la cadena k sea el resultado de la recombinación de las cadenas i y j . Ahora usando \mathcal{E} para denotar la esperanza tenemos que:

$$\mathcal{E} p_k^{t+1} = \sum_{ij} s_i^t s_j^t r_{ij}(k). \quad (2.6)$$

Para generalizar este modelo, la función $r_{ij}(k)$ es usada para construir una matriz de mezcla M donde la i, j -ésima entrada $m_{i,j} = r_{ij}(0)$. Notemos que esta matriz nos da las probabilidades de que cruzando las cadenas i y j produzcamos la cadena C_0 . En este caso asumimos que no hay mutación y que la cruza es de 1 punto. Entonces la matriz M es obviamente simétrica y cero en toda la diagonal salvo la entrada $m_{0,0}$ que es 1.0. Notemos que M es expresada completamente en términos de la información de ganancias. Por lo tanto, la primer columna/fila de la matriz es la inversa de la porción de las probabilidades de pérdidas de cadenas que corresponden a $\delta(C_i)$, donde cada cadena en el conjunto C_i es cruzado con C_0 . Por completitud, el valor de $\delta(C_i)$ para cadenas no producidas por los generadores de cadenas es 0.0 y, entonces la probabilidad de obtener S_0 durante la reproducción es 1.0. El resto de la matriz puede ser calculada usando los generadores de ganancias de cadenas y $\frac{\sigma(C_{\alpha+x}, C_{\omega+y})+1}{N-1}$, con $\sigma(C_{\alpha+x}, C_{\omega+y}) = N - \omega - \alpha$. Para cada par de cadenas producidas por los generadores de ganancias de cadenas determinar su índice e insertar el valor regresado por la función en la correspondiente localización en M . Por completitud, $\sigma(C_j, C_k) = -1$ para todos

los pares arbitrarios de cadenas no generados por los generadores de ganancias de cadenas, lo que implica que la entrada $m_{j,k} = 0$.

Una vez definida M no cambia dado que no es afectada por las variaciones de aptitud o por la representación proporcional en la población. Dadas las suposiciones: no hay mutaciones, que γ es actualizada cada generación para corregir los cambios en el promedio de la población y que se usa cruce de 1 punto, entonces la forma estándar de las ecuaciones ejecutables corresponde a la siguiente porción del modelo de Liepins y Vose [30]:

$$s^T M s$$

donde T denota la transposición. Mientras los términos de la aptitud son definidos explícitamente en la ecuación ejecutable, hasta ahora se ha asumido que s incluye la información sobre la aptitud sin indicar explícitamente como es calculada. La información de aptitud es agregada más adelante en el modelo de Vose y Liepins. Por ahora, sin embargo, esta expresión es primero generalizada para cubrir todas las cadenas en el espacio de búsqueda. Vose y Liepins formalizan la noción que el "or exclusivo" puede ser usado para remapear todas las cadenas en el espacio de búsqueda, en este caso representado por el vector s . Ellos muestran que si la recombinación es una combinación de cruce y mutación entonces

$$r_{i,j}(k \oplus q) = r_{i \oplus k, j \oplus k}(q) \text{ y específicamente } r_{i,j}(k \oplus 0) = r_{i \oplus k, j \oplus k}(0).$$

Esto nos permite reordenar los elementos en s con respecto a cualquier punto en particular en el espacio. Este reordenamiento es equivalente a un remapeo de variables en las ecuaciones ejecutables. Una función de permutación, ρ , está definida como:

$$\rho_j < s_0, \dots, s_{N-1} >^T = < s_{j \oplus 0}, \dots, s_{j \oplus (N-1)} >^T$$

donde los vectores son tratados como columnas y N es la dimensión del espacio de búsqueda. Ahora un operador general \mathcal{M} puede ser definido sobre s que remapea $s^T M s$ para cubrir todas las cadenas en el espacio de búsqueda.

$$\mathcal{M}(s) = < (\rho_0 s)^T M \rho_0 s, \dots, (\rho_{N-1} s)^T M \rho_{N-1} s >^T$$

Notemos que s tiene información que corresponde a la fase intermedia en el proceso de reproducción (después de la selección pero antes de la recombinación). Entonces para completar el ciclo y alcanzar el punto en el cual los modelos de Vose y Liepins se pueden ejecutar de forma iterativa, la información sobre la aptitud es ahora introducida para transformar la población al inicio de la iteración $t + 1$ a la siguiente población intermedia. Una matriz de aptitud F es definida de tal manera que la información sobre la aptitud es guardada en la diagonal; el elemento i , i -ésimo está dado por $f(i)$ donde f es la función de aptitud.

La transformación del vector p^{t+1} a la siguiente población intermedia representada por s^{t+1} está dada como sigue:

$$s^{t+1} \sim \mathcal{FM}(s^t).$$

Vose y Liepins dan ecuaciones para calcular la matriz de mezcla M la cual no solo incluye probabilidades para la cruza de 1 punto, sino también para la mutación. Una extensión más compleja del modelo de Vose y Liepins incluye modelos de poblaciones finitas usando modelos de Markov siguiendo el trabajo de Nix y Vose en [14]. Es decir, considera una población inicial, y la secuencia de poblaciones generadas a partir de ésta, donde cada población depende solamente de la anterior y no importa el orden de las cadenas en la población, es decir, poblaciones que difieren solo en permutaciones en las posiciones de los individuos son consideradas equivalentes. En este sentido, la óptica tomada en este documento es esta misma.

Sin embargo, uno de los problemas con esta formalización matemática es que no nos da una muy buena intuición en términos de los resultados clásicos. Otro problema potencial es que los resultados se presentan en términos de cadenas; sin embargo (como veremos adelante), se necesita una representación de más alto nivel para lidiar con la creciente complejidad que se presenta.

2.7. Cadenas de Markov

En esta sección presentamos una pequeña introducción, así como algunas definiciones y resultados importantes acerca de cadenas de Markov que usaremos en el capítulo 5 para obtener el comportamiento asintótico de las soluciones.

Describimos una cadena de Markov como sigue: tenemos un conjunto de estados, $S = \{s_1, s_2, \dots, s_r\}$. El proceso inicia en uno de estos estados y se mueve sucesivamente de un estado a otro. Cada movimiento se llama un *paso*. Si la cadena está actualmente en el estado s_i , entonces se mueve al estado s_j con una probabilidad denotada por p_{ij} , y esta probabilidad no depende de en cuales estados la cadena estuvo antes del actual.

Definición 2.7.1 *Llamaremos matriz de transición o matriz de transición de probabilidades a un arreglo cuadrado en el que las entradas son las probabilidades de transición entre estados.*

Teorema 2.7.1 *Sea P la matriz de transición de una cadena de Markov. La entrada ij -ésima $p_{ij}^{(n)}$ de la matriz P^n nos da la probabilidad de que la cadena de Markov, iniciando en el estado s_i , llegue al estado s_j después de n pasos.*

Ahora, dada la matriz de transición del sistema nos interesa saber cual será el estado del sistema después de iterarlo desde un estado inicial; la respuesta está en el siguiente teorema.

Teorema 2.7.2 *Sea P la matriz de transición de una cadena de Markov, y sea u el vector de probabilidad que representa la distribución inicial. Entonces la probabilidad de que la cadena esté en el estado s_i después de n pasos es la i -ésima entrada del vector*

$$u = uP^n$$

Definición 2.7.2 *Una cadena de Markov es llamada una cadena ergódica si es posible ir de cualquier estado a cualquier estado (no necesariamente en un paso). En algunos libros las cadenas de Markov ergódicas son llamadas irreducibles.*

Definición 2.7.3 *Una cadena de Markov es llamada una cadena regular si alguna potencia de la matriz de transición tiene solo elementos positivos.*

En otras palabras, para algún n es posible ir de cualquier estado a cualquier estado en exactamente n pasos. Es claro de la definición que cada cadena regular es ergódica. Por otro lado, una cadena ergódica no necesariamente es regular. Los AGs que empleamos son siempre ergódicos y en algunos casos (como veremos más adelante con mutación en el escenario plano y sin cruza) las cadenas son regulares.

Teorema 2.7.3 *Sea P la matriz de transición de una cadena regular. Entonces, si $n \rightarrow \infty$, las potencias de P^n se acercan a la matriz límite W con todas las filas iguales al vector w . El vector w es un vector de probabilidad estrictamente positivo (es decir, las entradas son todas positivas y suman a uno).*

Este teorema es de particular importancia para nuestro estudio pues generalmente nos interesa obtener la distribución límite de una cadena de Markov al ser iterada muchas veces (dada una configuración inicial), es decir, nos interesa conocer los puntos fijos del sistema en un escenario particular.

Capítulo 3

Dinámica de cuerdas

En las siguientes secciones presentamos una formulación de las ecuaciones para la dinámica de cuerdas inducida por los operadores genéticos más comunes: selección, mutación y cruce. Finalmente reunimos los resultados para obtener una ecuación general de la dinámica de cuerdas. En el resto del documento suponemos que estamos trabajando con una población infinita y por lo tanto las ecuaciones con las que trabajamos son exactas. Esta suposición obviamente no es cierta en la práctica, sin embargo lo que pretendemos es observar casos promedio y adquirir intuición.

3.1. Solo selección

La evolución, desde la óptica Darwiniana, esta regida primordialmente por el proceso de selección natural. En esta sección consideramos una población evolucionando en un contexto asexual y sin mutación, esto es de cierta manera una restricción fuerte, pero nos ayuda como punto de inicio para entender la dinámica de un AG. En términos más formales esto significa que $p_m = p_c = 0$. Entonces si consideramos selección proporcional

$$P(C_i, t + 1) = \frac{f(C_i)}{\bar{f}(t)} P(C_i, t)$$

donde $\bar{f}(t)$ es la aptitud promedio de la población al tiempo t . Notemos que usando este tipo de selección, las cuerdas con aptitud mayor que el promedio tienden a perdurar y multiplicarse para la siguiente generación.

En general, pero sobre todo en el caso natural, la función de aptitud depende del tiempo, sin embargo en el caso artificial por lo general nos restringimos a considerarla constante en el tiempo. Esta consideración es buena si suponemos que no hay demasiados cambios en el ambiente, o que dichos cambios se dan de manera gradual y suave.

Observamos que tenemos una solución formal inmediata que es:

$$P(C_i, t) = \prod_{j=0}^{t-1} \frac{f(C_i)}{f(j)} P(C_i, 0)$$

Esta solución parece no ser muy útil, dado que no podemos obtener resultados concretos por que debemos calcular $\bar{f}(j)$ directamente. Sin embargo observamos que cuando la cuerda C_i tiene un valor de aptitud ($f(C_i)$) mayor al del promedio ($\bar{f}(t)$) su proporción tiende a crecer, y viceversa. Si suponemos que $\frac{f(C_i)}{f(j)}$ es constante (α) en un periodo de tiempo $t - t'$ tenemos

$$P(C_i, t) = \alpha^{t-t'} P(C_i, t')$$

es decir el crecimiento o decrecimiento se presenta de manera "exponencial".

Siguiendo este esquema de selección proporcional nos encontramos con el teorema fundamental de selección natural de R. Fisher [1] que nos dice que la aptitud promedio siempre aumenta de una generación a otra, es decir:

Por definición de promedio tenemos:

$$\bar{f}(t+1) = \sum_{C_i} f(C_i) P(C_i, t+1) \quad (3.1)$$

luego sustituyendo P por lo que vale:

$$= \sum_{C_i} f(C_i) \frac{n(C_i, t+1)}{n} \quad (3.2)$$

Pero sabemos que la ecuación de evolución para el número de cadenas de tipo $C_i, n(C_i, t+1)$ es

$$n(C_i, t+1) = \frac{f(C_i, t)}{f(t)} n(C_i, t) \quad (3.3)$$

Así es que lo sustituimos (asumiendo que no hay dependencia temporal de $f(C_i)$) y tenemos:

$$= \sum_{C_i} f(C_i) \frac{f(C_i)}{f(t)} \frac{n(C_i, t)}{n} = \sum_{C_i} \frac{f(C_i)^2}{f(t)} \frac{n(C_i, t)}{n} = \sum_{C_i} \frac{f(C_i)^2}{f(t)} P(C_i, t) \quad (3.4)$$

por lo tanto

$$\bar{f}(t+1) - \bar{f}(t) = \sum_{C_i} \frac{f(C_i)^2}{f(t)} P(C_i, t) - \frac{1}{f(t)} \left(\sum_{C_i} f(C_i) P(C_i, t) \right)^2 = \left\langle \frac{(f - \bar{f})^2}{f} \right\rangle \geq 0 \quad (3.5)$$

es decir, la aptitud nunca decrece. Podemos hacer una prueba para el caso continuo, la demostración procede tomando la derivada del promedio de la función de aptitud.

Donde en el límite cuando tenemos una población infinita, $P(C_i, t)$ tiende a ser la probabilidad de encontrar la cadena C_i al tiempo t ; en el caso de una

población finita tendremos algunas variaciones debidas a los procesos estocásticos a las que llamamos efecto de tamaño finito. Para calcular la evolución del promedio de la función de aptitud como función del tiempo necesitamos resolver un sistema de 2^n ecuaciones simultáneas.

3.2. Selección y mutación

El operador de mutación es la manera natural de introducir variaciones en los nuevos individuos de una población, en términos de búsqueda estamos hablando de exploración del espacio. En esta sección se muestra la ecuación de evolución para una población en la que no hay intercambio genético entre los individuos, en otras palabras la probabilidad de cruza es cero ($p_c = 0$).

Para la ecuación de evolución de la población considerando mutación debemos considerar que en general para N bits la proporción esperada de la cadena C_i al tiempo $t + 1$ es su proporción al tiempo t por la probabilidad de que ninguno de sus bits cambie más la probabilidad de que alguna cadena C_j se convierta en C_i multiplicado por la probabilidad de que C_j exista en la población. Esta última probabilidad está compuesta por la probabilidad de que los bits en los que difiere C_j cambien y la probabilidad de que en los que tienen en común no ocurra ningún cambio. Por supuesto que entre más "diferentes" sean las cadenas hay menor probabilidad de que una se mute en la otra, es decir:

$$\mathcal{P}(C_i, t + 1) = \mathcal{P}(C_i \rightarrow C_i)P'(C_i, t) + \sum_{C_j \neq C_i} \mathcal{P}(C_j \rightarrow C_i)P'(C_j, t) \quad (3.6)$$

donde $\mathcal{P}(C_i \rightarrow C_i) = (1 - p_m)^N$ es la probabilidad de que la cadena C_i se mantenga sin cambios debidos a la mutación; ésto es que ninguno de sus N bits cambie, asumiendo la misma probabilidad de mutación para cada uno y

$$\mathcal{P}(C_j \rightarrow C_i) = (1 - p_m)^{N - d^H(i,j)} p_m^{d^H(i,j)} \quad (3.7)$$

es la probabilidad de que la cadena C_j sea mutada a C_i ; esto es, que los bits que tienen en común se mantengan intactos y los bits en los que difieren en C_j cambien, siendo $d^H(i, j)$ la distancia de Hamming entre la cadena C_i y la cadena C_j .

La siguiente generalización es considerar el caso en que p_m depende de la posición del bit en la cadena, donde la probabilidad de que la cadena C_j sea mutada a C_i está dada por

$$\mathcal{P}(C_j \rightarrow C_i) = \prod_{k \in \{C_j - C_i\}} p_m(k) \prod_{k \in \{C_j - C_i\}^c} (1 - p_m(k)) \quad (3.8)$$

En donde expresamos la dependencia de la probabilidad de mutación en la posición explícitamente en el término $p_m(k)$ que es la probabilidad de mutación del bit k y $\{C_j - C_i\}$ es el conjunto de bits en que difieren C_j y C_i y $\{C_j - C_i\}^c$ es el complemento de este conjunto, es decir los bits en los que coinciden.

Una forma simplificada de expresar la ecuación anterior es mediante la notación matricial

$$\mathbf{P}(t+1) = \overline{\mathbf{W}}\mathbf{P}'(t) \quad (3.9)$$

Donde: $\mathbf{P}(t+1)$ es el vector cuyas entradas son las proporciones de cada una de las posibles cadenas de la población al tiempo $t+1$, $\mathbf{P}'(t)$ es el vector de proporciones después de la selección al tiempo t y la matriz de mutación $\overline{\mathbf{W}}$ es real, simétrica e independiente del tiempo y sus elementos son $\overline{W}_{ij} = (1 - p_m)^{N-d''(i,j)} p_m^{d''(i,j)}$.

Igual que antes podemos escribir la solución formal:

$$\mathbf{P}(t) = \prod_{j=0}^{t-1} \overline{\mathbf{W}}_s(j) \mathbf{P}(0)$$

Aquí $\overline{\mathbf{W}}_s(j)$ es la matriz que representa la agrupación de selección y mutación al tiempo j . Es decir, es la matriz $\overline{\mathbf{W}}$ multiplicada por la matriz diagonal $\text{diag}\left\{\frac{f(C_0)}{f(j)}, \dots, \frac{f(C_{n-1})}{f(j)}\right\}$. Ahora la matriz $\overline{\mathbf{W}}_s(j)$ tiene una dependencia en el tiempo dada por la función de aptitud promedio.

3.3. Selección y cruza

El apareamiento en la naturaleza tiene el fin de hacer intercambio de información genética entre los individuos de la población. Dicho apareamiento se hace sobre individuos que han sido seleccionados de manera aleatoria entre los más aptos y algunos que probablemente no lo sean, dado que se pueden obtener características buenas de individuos que no destacan.

En esta sección exploramos la ecuación de evolución para el operador de cruza después de un mecanismo de selección. Para este estudio consideramos inicialmente cruza de 1 punto, y después presentamos la ecuación para cruza general con máscaras. A continuación consideramos la proporción de cadenas del tipo C_i en la población, después de la selección y cruza.

El número de cadenas de tipo C_i después de la cruza es el número de cadenas del tipo C_i después de la selección sumándole el número de cadenas construidas a partir de otras $C_j, C_l \neq C_i$ reputándole el número de cadenas que fueron destruidas por la cruza. Podemos ver esto como: (cadenas después de selección) - (cadenas destruidas en la cruza) + (cadenas construidas con la cruza).

Observemos el término de (cadenas destruidas en la cruza), tenemos que considerar los siguientes hechos:

- Suponemos que tenemos p_c como probabilidad de cruza.
- Entonces C_i es destruido al cruzarse con $C_j \Leftrightarrow C_j$ es "diferente" de C_i tanto a la izquierda como a la derecha del punto de cruza. Una manera de caracterizar estas diferencias es a través de la distancia de Hamming, es

decir 2 cadenas son diferentes \Leftrightarrow su distancia de Hamming es positiva $\Leftrightarrow d_L^H(i, j) > 0$ y $d_R^H(i, j) > 0$, donde $d_L^H(i, j)$ (alt. $d_R^H(i, j)$) es la distancia de Hamming entre las partes a la izquierda (alt. derecha) de las cadenas i y j con respecto al punto de cruce, entonces introducimos el factor:

$$\mathcal{C}_{C_i, C_j}^{(1)} = \theta(d_L^H(i, j))\theta(d_R^H(i, j)) \quad (3.10)$$

Donde $\theta(x) = 1$ si $x > 0$ y $\theta(0) = 0$

- C_i y C_j tienen por probabilidad de cruce $P'(C_i, t)P'(C_j, t)$ para cada $C_i \neq C_j$. Ahora tenemos que tomar un promedio sobre todos los posibles puntos de cruce $\frac{1}{N-1} \sum_{k=1}^{N-1}$

Por lo tanto (cadenas destruidas en la cruce)=

$$\begin{aligned} p_c \sum_{C_i \neq C_j} \frac{1}{N-1} \sum_{k=1}^{N-1} \mathcal{C}_{C_i, C_j}^{(1)} P'(C_i, t) P'(C_j, t) \\ = \frac{p_c}{N-1} \sum_{C_i \neq C_j} \sum_{k=1}^{N-1} \mathcal{C}_{C_i, C_j}^{(1)} P'(C_i, t) P'(C_j, t) \end{aligned} \quad (3.11)$$

Ahora tenemos que considerar las cadenas construidas en la cruce. La lógica es la siguiente:

C_i es construida a partir de las cadenas C_j y C_l si y solo si C_j es idéntica a la izquierda con C_i y C_l es idéntica a la derecha con C_i o viceversa. Si y solo si $(C_i^R = C_j^R \text{ y } C_i^L = C_l^L)$ o $(C_i^R = C_l^R \text{ y } C_i^L = C_j^L)$ si y solo si $(d_R^H(i, j) = 0 \text{ y } d_L^H(i, l) = 0)$ o $(d_R^H(i, l) = 0 \text{ y } d_L^H(i, j) = 0)$ si y solo si $\delta(d_R^H(i, j))\delta(d_L^H(i, l)) + \delta(d_R^H(i, l))\delta(d_L^H(i, j)) > 0$. Por lo tanto definimos el término:

$$\mathcal{C}_{C_i, C_j}^{(2)} = \frac{1}{2} [\delta(d_R^H(i, j))\delta(d_L^H(i, l)) + \delta(d_R^H(i, l))\delta(d_L^H(i, j))] \quad (3.12)$$

Con $\delta(x) = 0 \forall x \neq 0$ y $\delta(0) = 1$. Y podemos decir que la proporción de cadenas construidas en la cruce es:

$$\frac{p_c}{N-1} \sum_{C_j \neq C_i} \sum_{C_l \neq C_i} \sum_{k=1}^{N-1} \mathcal{C}_{C_i, C_j}^{(2)} P'(C_j, t) P'(C_l, t) \quad (3.13)$$

Ahora solo nos resta unir las ecuaciones recientemente derivadas para obtener

$$\begin{aligned} P(C_i, t+1) = P'(C_i, t) - \frac{p_c}{N-1} \sum_{C_i \neq C_j} \sum_{k=1}^{N-1} \mathcal{C}_{C_i, C_j}^{(1)} P'(C_i, t) P'(C_j, t) \\ + \frac{p_c}{N-1} \sum_{C_j \neq C_i} \sum_{C_l \neq C_i} \sum_{k=1}^{N-1} \mathcal{C}_{C_i, C_j}^{(2)} P'(C_j, t) P'(C_l, t) \end{aligned} \quad (3.14)$$

Podemos modelar la cruza en términos generales mediante máscaras como sigue:

$$P_c(C_i, t) = P'(C_i, t) - \sum_{m=1}^{2^N} p_c(m) \sum_{C_i \neq C_j} C_{C_i, C_j}^{(1)}(m) P'(C_i, t) P'(C_i, t) \\ + \sum_{m=1}^{2^N} p_c(m) \sum_{C_i \neq C_j} \sum_{C_l \neq C_i} C_{C_j, C_l}^{(2)}(m) P'(C_j, t) P'(C_l, t) \quad (3.15)$$

donde $\sum_{m=1}^{2^N}$ es la suma sobre todas las máscaras posibles de cruza $m \in \mathcal{M}$, siendo \mathcal{M} el espacio de máscaras y $p_c(m)$ la probabilidad de implementar la máscara m . Para una máscara en general, denotaremos por s los bits que son tomados de un padre y c para los del otro. Entonces

$$C_{C_i, C_j}^{(1)}(m) = \theta(d_s^H(i, j)) \theta(d_c^H(i, j)) \quad (3.16)$$

y

$$C_{C_i, C_j}^{(2)}(m) = \frac{1}{2} [\delta(d_s^H(i, j)) \delta(d_c^H(i, l)) + \delta(d_c^H(i, l)) \delta(d_s^H(i, j))] \quad (3.17)$$

donde $d_s^H(i, j)$ es la distancia de Hamming entre las cadenas C_i y C_j medida solo sobre los bits del conjunto s , con los otros argumentos definidos como antes. Podemos considerar la cruza de un punto como un caso particular. En este caso hay solo $N-1$ máscaras de cruza etiquetadas por el punto k . Entonces, $p_c(m) = p_c/(N-1)$ para $m = k, k \in [1, N-1]$ y $p_c(m) = 0$ de otra manera. También $s = L$ y $c = R$ (o viceversa) donde L y R se refieren a la parte izquierda y derecha del punto de cruza respectivamente. Para el caso de cruza de 2 puntos hay $N-1 C_2$ máscaras no cero etiquetadas por los dos puntos de cruza, k_1 y k_2 . Entonces $p_c(m) = p_c/N-1 C_2$ para $m \in \{k_1, k_2\}$ con $k_1, k_2 \in [1, N-1]$ y $k_1 > k_2$. En este caso s representa las partes "fuera" de k_1 y k_2 y c la parte entre ellos (o viceversa). Como último ejemplo, para cruza uniforme $p_c(m) = p_c p_u^{N_s} (1 - p_u)^{N - N_s} / 2^N$ donde p_u es la probabilidad de que un allele dado sea heredado de C_j .

A continuación presentamos la derivación de una gran simplificación para las ecuaciones que involucran la cruza. Reconsideremos la ecuación (3.14). El segundo término del lado derecho de (3.14) es cero salvo cuando C_j difiere en al menos un bit tanto a la izquierda como a la derecha del punto de cruza. Ahora podemos escribir:

$$\sum_{C_j \neq C_i} C_{ij}^{(1)}(k) P'(C_j, t) = \left(\sum_{C_j} - \sum_{C_j, C_j^l, C_j^r} - \delta_{C_i, C_j} \right) P'(C_j, t) \quad (3.18)$$

donde \sum_{C_j, C_j^l, C_j^r} es la suma sobre todas las cadenas que no difieren en ningún bit de C_i a la izquierda y cualquier número a la derecha o ninguno a la derecha y cualquiera a la izquierda y δ_{C_i, C_j} es 1 si $C_i = C_j$ y 0 de otra manera. Dado que $\sum_{C_j} P(C_j, t) = 1$ tenemos:

$$\begin{aligned}
P(C_i, t+1) &= P'(C_i, t) - \frac{p_c}{N-1} \sum_{k=1}^{N-1} \left(\sum_{C_j} P'(C_i, t) P'(C_j, t) \right. \\
&\quad \left. - P'(C_i, t) \sum_{C_j, C_i^L, C_i^R} P'(C_j, t) - P'^2(C_i, t) \right) \\
&\quad + \frac{p_c}{N-1} \sum_{C_j \neq C_i} \sum_{C_l \neq C_i} \sum_{k=1}^{N-1} C_{C_i, C_j}^{(2)} P'(C_j, t) P'(C_l, t) \\
&= P'(C_i, t)(1 - p_c) + \frac{p_c}{N-1} \sum_{k=1}^{N-1} \sum_{C_j, C_i^L, C_i^R} P'(C_i, t) P'(C_j, t) + p_c P'(C_i, t) P'(C_i, t) \\
&\quad + \frac{p_c}{N-1} \sum_{C_j \neq C_i} \sum_{C_l \neq C_i} \sum_{k=1}^{N-1} C_{C_i, C_j}^{(2)} P'(C_j, t) P'(C_l, t)
\end{aligned} \tag{3.19}$$

Ahora considerando el último término y dado que las sumas son sobre C_j y $C_l \neq C_i$, es decir, cadenas que difieren en al menos un bit de C_i . $C_{C_i, C_j}^{(2)}$ significa que C_j difiere en al menos un bit a un lado de k pero no en el otro. Similarmente para C_l .

Ahora

$$\sum_{C_j} \sum_{C_l} = \sum_{\substack{C_j \neq C_i \\ C_l \neq C_i}} + \sum_{\substack{C_j = C_i \\ C_l \neq C_i}} + \sum_{\substack{C_j \neq C_i \\ C_l = C_i}} + 1 \tag{3.20}$$

donde 1 representa el término donde ambos C_j y C_l son C_i . Por lo tanto el último término en (3.19) es:

$$\begin{aligned}
&\frac{p_c}{N-1} \sum_{k=1}^{N-1} \left(\sum_{C_j} \sum_{C_l} C_{C_i, C_j}^{(2)} P'(C_j, t) P'(C_l, t) - \sum_{\substack{C_j = C_i \\ C_l \neq C_i}} C_{C_i, C_j}^{(2)} P'(C_j, t) P'(C_l, t) \right. \\
&\quad \left. - \sum_{\substack{C_j \neq C_i \\ C_l = C_i}} C_{C_i, C_j}^{(2)} P'(C_j, t) P'(C_l, t) - \sum_{C_j, C_l} \delta_{C_j, C_i} \delta_{C_l, C_i} P'(C_l, t) P'(C_j, t) \right)
\end{aligned} \tag{3.21}$$

El segundo término es:

$$\begin{aligned}
&\sum_{C_j, C_l} \delta_{C_j, C_i} \frac{1}{2} (\delta(d_L^H(C_i, C_j)) (\delta(d_R^H(C_i, C_l)) + \overleftarrow{L R})) P'(C_l, t) P'(C_j, t) = \\
&\sum_{C_l} \frac{1}{2} (\delta(d_L^H(C_i, C_l)) \delta(d_R^H(C_i, C_l)) + \delta(d_L^H(C_i, C_l)) \delta(d_R^H(C_i, C_l))) P'(C_l, t) P'(C_j, t) \\
&= \frac{1}{2} \sum_{C_l} (\delta(d_R^H(C_i, C_l)) + \delta(d_L^H(C_i, C_l))) P'(C_l, t) P'(C_j, t)
\end{aligned} \tag{3.22}$$

aquí \overleftarrow{LR} se refiere a intercambiar los subíndices R por L en el término anterior y viceversa. El tercer término es lo mismo pero con $C_l = C_i$ y $C_j \neq C_i$; pero estas son solo etiquetas, así que podemos cambiar l por j en (3.22) y cancelar el $\frac{1}{2}$ para obtener

$$\sum_{C_j \neq C_i} (\delta(d_R^H(C_i, C_j)) + \delta(d_L^H(C_i, C_j))) P'(C_l, t) P'(C_j, t) \quad (3.23)$$

El último término en (3.21) es simplemente $P'(C_i, t) P'(C_i, t)$ que se cancela con el tercer término de (3.19), así

$$\begin{aligned} P(C_i, t+1) &= P'(C_i, t)(1-p_c) + \frac{p_c}{N-1} \sum_{k=1}^{N-1} \sum_{C_j, C_i^k, C_i^k} P'(C_i, t) P'(C_j, t) \\ &\quad + \frac{p_c}{N-1} \sum_{C_j \neq C_i} \sum_{C_l \neq C_i} \sum_{k=1}^{N-1} C_{C_i, C_j}^{(2)} P'(C_j, t) P'(C_l, t) \\ &\quad - \sum_{C_j \neq C_i} (\delta(d_R^H(C_i, C_j)) + \delta(d_L^H(C_i, C_j))) P'(C_i, t) P'(C_j, t) \end{aligned} \quad (3.24)$$

Entonces el segundo término es sobre los C_j que tienen C_i^L pero difieren en al menos un bit a la derecha o tienen C_i^R y difieren en al menos un bit a la izquierda; pero esto puede ser escrito como:

$$\sum_{C_j \neq C_i} (\delta(d_R^H(C_i, C_j)) + \delta(d_L^H(C_i, C_j))) P'(C_i, t) P'(C_j, t) \quad (3.25)$$

Entonces podemos ver que el segundo y el cuarto término se cancelan exactamente

$$P(C_i, t+1) = P'(C_i, t)(1-p_c) + \frac{p_c}{N-1} \sum_{C_j} \sum_{C_l} \sum_{k=1}^{N-1} C_{C_i, C_j}^{(2)} P'(C_j, t) P'(C_l, t) \quad (3.26)$$

Ahora podemos cambiar el segundo término considerando que

$$\sum_{C_j} \sum_{C_l} C_{C_i, C_j}^{(2)} = \frac{1}{2} \left(\sum_{C_j \in C_i^L} \sum_{C_l \in C_i^R} + \sum_{C_j \in C_i^R} \sum_{C_l \in C_i^L} \right) \quad (3.27)$$

Ahora

$$\sum_{C_j \in C_i^L} P'(C_j, t) = P'(C_i^L, t); \quad \sum_{C_l \in C_i^R} P'(C_l, t) = P'(C_i^R, t) \quad (3.28)$$

Donde $P'(C_i^L, t) = (\bar{f}(C_i^L)/\bar{f})P(C_i^L, t)$ y $\bar{f}(C_i^L)$ es el promedio de la función de aptitud evaluada en las cadenas que contienen C_i^L (la parte izquierda de

C_i respecto al punto de cruce). Aquí vemos como la noción de esquemas se introduce al nivel de cadenas. Entonces:

$$P(C_i, t + 1) = P'(C_i, t)(1 - p_c) + \frac{p_c}{N - 1} \sum_{k=1}^{N-1} P'(C_i^L(k), t)P'(C_i^R(k), t) \quad (3.29)$$

Donde izquierda (L) y derecha (R) son respecto al punto de cruce (k).

3.4. Representación de “bloques constructores”

Podemos resumir la derivación de la sección anterior. Reconsideremos la ecuación (3.14). En el término de destrucción restringimos la suma a aquellos C_j que difieren de C_i en al menos un bit a ambos lados del punto de cruce. Se puede convertir la suma sobre los C_j a una suma sin restricciones substrayendo aquellos C_j que tienen $d_L^H(i, j) = 0$ y/o $d_R^H(i, j) = 0$. Similarmente podemos escribir el término de reconstrucción como [26]:

$$\sum_{C_j \in C_i^L} \sum_{C_l \in C_i^R} P'(C_j, t)P'(C_l, t) \quad (3.30)$$

donde C_i^L es la parte de C_i a la izquierda del punto de cruce y análogamente C_i^R . Sin embargo,

$$\bar{f}(C_i^L, t) = \frac{1}{n_{C_i^L}} \sum_{C_j \in C_i^L} P'(C_j, t) \quad (3.31)$$

donde $n_{C_i^L}$ es el número de cadenas en la población que contienen C_i^L .

La gran simplificación que observamos nos sugiere la utilización de algún mecanismo para la división de las cadenas y poder trabajar con subcadenas. La herramienta que tradicionalmente se utiliza son los “esquemas” que estudiaremos con mayor detalle en el siguiente capítulo. En esencia la hipótesis de bloques constructores dice que un AG llega a una solución óptima de un problema complejo vía la combinación de subcadenas pequeñas con valor alto de aptitud. Sin embargo, no hay nada que indique que deban ser pequeñas, de hecho se pueden construir ejemplos en lo que ocurre lo contrario. Esta forma de dividir las cadenas en izquierda y derecha del punto de cruce nos sugiere la aparición de estos bloques cada vez más pequeños que se acoplan sucesivamente en el tiempo hasta llegar a la solución final.

3.5. Ecuación general microscópica

En las secciones anteriores hemos considerado los tres operadores genéticos más comunes: selección, mutación y cruce por separado. También lo podemos ver como que restringimos la ecuación general a estos casos simples, y dimos

seguimiento a la derivación de cada fórmula en particular. Dichas restricciones nos permiten tener una mejor óptica de lo que ocurre en el escenario general. Ahora consideraremos la ecuación general para las proporciones de las cadenas. Esta ecuación es básicamente el ensamble de las anteriores.

$$P(C_i, t + 1) = P(C_i \rightarrow C_i)P_c(C_i, t) + \sum_{C_j \neq C_i} P(C_j \rightarrow C_i)P_c(C_j, t) \quad (3.32)$$

Donde $P_c(C_i, t)$ representa las proporciones de la cadena C_i después del cruce y selección, es decir:

$$\begin{aligned} P_c(C_i, t) = P'(C_i, t) - \frac{p_c}{N-1} \sum_{C_i \neq C_j} \sum_{k=1}^{N-1} C_{C_i, C_j}^{(1)} P'(C_i, t) P'(C_j, t) \\ + \frac{p_c}{N-1} \sum_{C_j \neq C_i} \sum_{C_l \neq C_i} \sum_{k=1}^{N-1} C_{C_i, C_j}^{(2)'} P'(C_j, t) P'(C_l, t) \end{aligned} \quad (3.33)$$

o en su forma simplificada:

$$P_c(C_i, t) = P'(C_i, t)(1 - p_c) + \frac{p_c}{N-1} \sum_{k=1}^{N-1} P'(C_i^L(k), t) P'(C_i^R(k), t) \quad (3.34)$$

que extendida a cualquier cruce homóloga tiene la siguiente forma:

$$P_c(C_i, t) = P'(C_i, t)(1 - p_c) + \sum_{m=0}^{2^N} p_c(m) P'(C_i^s(m), t) P'(C_i^c(m), t) \quad (3.35)$$

Notemos que para obtener una solución a la ecuación anterior necesitamos resolver 2^N ecuaciones en diferencias no lineales y acopladas (o diferenciales en el caso continuo). Notemos también que la ecuación (3.14) es equivalente a los resultados planteados por Vose [30].

Podemos expresar los resultados obtenidos en la sección 3.3 en notación matricial como sigue:

$$\mathbf{P}_c(t) = \bar{\mathbf{F}}(t)\mathbf{P}(t) + \sum_{m=1}^{2^N} p_c(m)\mathbf{j}(m, t) \quad (3.36)$$

que es una ecuación lineal con fuente, donde la matriz de selección y destrucción por cruce, $\bar{\mathbf{F}}(t)$, es diagonal. El vector $\mathbf{j}(m, t)$ tiene por componentes a: $P'(C_i^s(m), t)P'(C_i^c(m), t)$ y definiendo la matriz de selección, destrucción por cruce y mutación $\bar{\mathbf{W}}_s(t) = \bar{\mathbf{W}}\bar{\mathbf{F}}(t)$ tenemos

$$\mathbf{P}(t + 1) = \bar{\mathbf{W}}_s(t)\mathbf{P}(t) + \sum_{m=1}^{2^N} p_c(m)\bar{\mathbf{W}}\mathbf{j}(m, t) \quad (3.37)$$

La interpretación de esta ecuación es que $\mathbf{j}(m, t)$ es una “fuente” que crea cadenas (o esquemas) juntando bloques constructores, mientras el primer término del lado derecho nos dice como las cadenas son propagadas a la siguiente generación. El efecto destructivo de la cruza renormaliza la aptitud de las cadenas. La solución iterada a la ecuación (3.36) es, como se muestra en [23]:

$$\mathbf{P}(t) = \prod_{n=0}^{t-1} (\overline{\mathbf{W}}_s(n)) \mathbf{P}(0) + \sum_{m=1}^{2^N} p_c(m) \sum_{n=0}^{t-1} \prod_{i=n}^{t-1} \overline{\mathbf{W}}_s(i) \overline{\mathbf{W}} \mathbf{j}(n)$$

Cuando Holland planteó su teorema de esquemas a mediados de los 70s lo hizo de manera simplificada considerando únicamente el efecto destructivo de la cruza [11]. La hipótesis de bloques constructores asegura que los bloques cortos y aptos son favorecidos. En [28, 26], sin embargo, se muestra que cuando el escenario de aptitud favorece la reconstrucción, los bloques más grandes prevalecen. El tener una ecuación en términos exactos nos permite definir una medida de aptitud más exacta que, a diferencia de la función de aptitud tradicional, considere estos efectos [24, 25]

Capítulo 4

Cambio de escala y dinámica de grados de libertad efectivos

Una de las principales dificultades que aparecen al resolver una gran variedad de problemas (aun los de apariencia sencilla) es la gran cantidad de grados de libertad que surgen. Usualmente entre más detalle se considere, las ecuaciones para describir el sistema crecen tanto en complejidad como en número; dicha complejidad puede convertir el problema en prácticamente irresoluble. Un método empleado en la física estadística para atacar el problema de crecimiento de complejidad es el cambio de escala, que consiste en ver al sistema con un menor nivel de detalle, ignorando los detalles irrelevantes para la resolución del problema. Esto nos permite tener menos ecuaciones a cambio de perder precisión. Un ejemplo de la aplicación de este método es el siguiente: en la genética poblacional es común el simular casos de 1 y 2 genes. Esta simplificación resulta natural en los siguientes casos:

1. Los genes están altamente correlacionados y podemos verlos simplemente como 2 grandes bloques.
2. Solo nos interesa ver el comportamiento de esos 2 genes independientemente de los demás, asumiendo poca epistasis con otros genes.
3. Esta simplificación nos permite resolver explícitamente algunas ecuaciones del comportamiento de los genes.

Esto simplifica en gran medida el estudio pues reduce considerablemente el tamaño del espacio de búsqueda, y aunque solo nos da una aproximación al resultado real, asumimos que dadas estas condiciones la aproximación es buena.

Existen dos tipos de cambio de escala: aproximado y exacto. En el primero ignoramos variables que asumimos son de poca importancia para la resolución

del problema o están altamente correlacionas con otras que mantendremos. En el cambio de escala exacto buscamos transformaciones del espacio de búsqueda que nos permitan observar las simetrías que se traducen en simplificaciones de las ecuaciones de evolución. Los cambios de escala son comunes en la física estadística en donde con frecuencia se estudian sistemas con $\sim 10^{23}$ de grados de libertad. En este documento optaremos por cambios de escala exactos dado que las aplicaciones se restringen a modelos de juguete en sistemas en los que observamos una simetría clara del espacio de búsqueda.

4.1. Elementos generales

Sea \mathcal{G} el espacio de genotipos y sea \mathcal{CG} una relación que induce una cubierta sobre \mathcal{G} . Nos interesa observar el comportamiento de cada uno de los elementos de la cubierta por separado, esperando que esto nos esboce el comportamiento de los elementos de \mathcal{G} . Claramente hay muchos cambios de escala para un conjunto dado; sin embargo no todos son “naturales” o reflejan la naturaleza de los grados de libertad efectivos, es decir, no todos mantienen una covarianza en la forma de las ecuaciones o llevan a simplificaciones de las mismas. Dicha covarianza es deseable pues nos permite tener una intuición clara sobre la evolución de la población a diferentes resoluciones. En otras palabras, la idea básica de hacer un cambio de escala es moverse a un espacio de estados reducido. En nuestro estudio en particular nos interesa mantener una regularidad en las expresiones para aprovechar las soluciones en el caso de cuerdas. Para la formalización del concepto de cambio de escala, iniciamos definiendo un observable como:

$$\langle o(t) \rangle = \sum_{C_i} o(C_i, t) P(C_i, t)$$

Un observable representa una propiedad para muchos grados de libertad (cadenas) con características comunes. Un ejemplo de observable es la aptitud promedio, definido por:

$$\bar{f}(t) = \sum_{C_i} f(C_i, t) P(C_i, t)$$

En el caso particular de la aptitud promedio, es una característica que requiere de toda la población; sin embargo existen algunos otros observables (como veremos más adelante) que solo requieren de las cadenas contenidas en un subconjunto de la población.

Podemos definir cambios de escala, incluso una sucesión de ellos en la que se realicen particiones sucesivas del conjunto, iniciando en una partición fina (considerando cada cuerda por separado) hasta la escala más grande en la que trata a toda la población como un solo conjunto. Como es usual los casos de interés se encuentran en un punto intermedio y rara vez en los extremos. En la figura (4.1) se muestra de manera esquemática una sucesión de cambios de escala, se ve como a medida que aplicamos cambios de escala (nos movemos

hacia la derecha) reducimos los grados de libertad con la consecuente pérdida de resolución.

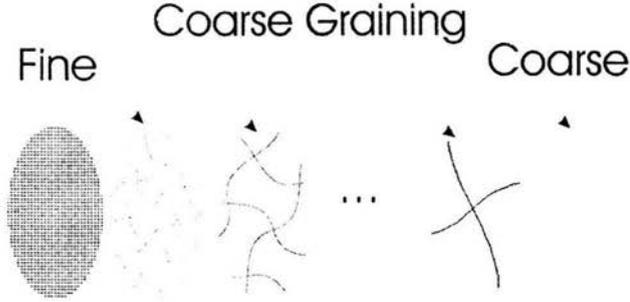


Figura 4.1: Diagrama con diferentes cambios de escala, iniciando por las cuerdas a la izquierda y terminando por el máximo nivel en el que la cubierta es el mismo conjunto.

El efecto de hacer un cambio de escala sobre un espacio de genotipos es el de un cambio de base que nos permite tener una óptica global (a cierto nivel), sin perder las propiedades importantes de las cadenas. Esta visión escénica nos da la oportunidad de trabajar sobre diferentes agrupaciones de cadenas, considerando cada elemento de la cubierta como un ente por separado.

Sea Γ_j un elemento de la cubierta de un espacio con un cambio de escala CG inducido por algún cambio de base o proyección. Usaremos la siguiente notación para el paso de variables sobre cadenas a variables escaladas:

$$F_{CG}(\Gamma_j) = \sum_{C_i \in \Gamma_j} F(C_i) \quad (4.1)$$

Con F una función definida para cadenas. Con frecuencia confundiremos intencionalmente F_{CG} con F en los casos en los que quede clara la diferencia. Equivalentemente usaremos la siguiente notación:

$$\mathcal{R}_{CG}^{\Gamma_j} F(C_i) = F_{CG}(\Gamma_j) \quad (4.2)$$

Llamaremos a $\mathcal{R}_{CG}^{\Gamma_j}$ operador de cambio de escala (o de renormalización). De manera natural podemos dotar al conjunto de dichos operadores de una estructura de semigrupo mediante la operación de composición.

$$\mathcal{R}_{CG}^{\Gamma_i} F(C_i) = \mathcal{R}_{CG}^{\Gamma_j} \mathcal{R}_{CG}^{\Gamma_k} F(C_i) \quad (4.3)$$

La identidad en el semigrupo es el operador definido por la relación identidad, es decir, $x \sim y$ si y solo si $x = y$. Claramente el conjunto de operadores de cambio de escala con la operación de composición no tiene inversos; lo podemos ver intuitivamente pues un cambio de escala implica perder resolución. En términos

menos formales, la ecuación anterior nos dice que podemos observar la dinámica de los diferentes componentes (a diferentes niveles de detalle). Esto crea una estructura de árbol en el que las hojas son las cadenas individuales y la raíz es la población completa considerada como un solo ente; cada aplicación de un operador de cambio de escala nos lleva un nivel arriba en la jerarquía inducida. Esto implica una reducción en el número de ecuaciones aunque puede hacerlas más complejas. Un ejemplo de la aplicación de estos operadores es el siguiente:

$$\mathcal{R}_{CG}^{\Gamma_i} P(C_i, t) = P_{CG}(\Gamma_j, t) = \sum_{C_i \in \Gamma_j} P(C_i, t) = P(\Gamma_j, t) \quad (4.4)$$

En el caso de proporciones, notemos que a pesar de que P es función de C_i y de t , el cambio de escala se hace solo sobre la variable que se refiere a las cadenas. Otro ejemplo es para el caso de la función de aptitud, en el cual tenemos:

$$f_{CG}(\Gamma_j) = \frac{1}{|\Gamma_j|} \sum_{C_i \in \Gamma_j} f(C_i) \quad (4.5)$$

Debemos notar que en este caso es necesario el término de normalización adicional $\frac{1}{|\Gamma_j|}$ donde $|\cdot|$ denota la cardinalidad del conjunto.

4.2. Operadores genéticos escalados

Sea \mathcal{H} un operador genético en general como en (2.3). Nuestro objetivo ahora es ver el comportamiento de \mathcal{H} una vez aplicado el operador de cambio de escala \mathcal{R}_{CG} . Es decir, dada la ecuación (2.2) tenemos:

$$\mathcal{R}_{CG} \mathbf{P}(C_i, t + 1) = \mathcal{R}_{CG} \mathcal{H} \mathbf{P}(C_i, t) \quad (4.6)$$

Y queremos saber si existe \mathcal{R}_{CG} tal que podamos observar \mathcal{H}_{Γ_j} ,

$$\mathbf{P}(\Gamma_j, t + 1) = \mathcal{H}_{\Gamma_j} \mathbf{P}(\Gamma_j, t) \quad (4.7)$$

Es decir, si podemos elegir un cambio de escala que mantenga una covarianza de forma en las ecuaciones de evolución. Debemos notar que si bien podemos definir una gran cantidad de cambio de escala no todos son “naturales”. Por ejemplo si queremos hacer un estudio sobre la anatomía del hombre, un cambio de escala es dividir el cuerpo en cubos de cierto tamaño. Sin embargo, esto no nos dará una visión que nos ayude a resolver el problema (entender la anatomía) con facilidad. Un cambio de escala más “natural” en este sentido sería por ejemplo estudiar la anatomía del cuerpo humano mediante una división en órganos. Siguiendo esta lógica, y como veremos a continuación, los esquemas son un coarse graining natural y exacto con respecto al operador de cruza; sin embargo, algunas otras descripciones no exactas pueden ser de utilidad también.

4.3. Esquemas

Durante mucho tiempo la herramienta primordial para el estudio de los AGs ha sido el uso de esquemas (plantillas). Estos nos permiten considerar el comportamiento de ciertos genes sin tomar en cuenta otros, es decir nos dan intuitivamente la noción de subproblemas. A continuación mostraremos la formalización de este concepto.

Consideremos el alfabeto Ω sobre el que opera el AG, definimos el conjunto de esquemas Ω' de longitud N sobre Ω como: $(\Omega \cup \{*\})^N$. En otros términos, anexamos el símbolo “*” a nuestro alfabeto, y por lo tanto extendemos nuestro espacio de búsqueda. A “*” lo conocemos como el símbolo “don’t care” o “no importa”. Algunos ejemplos de esquemas para $N = 11$ son: $\xi_0 = 01**10*110*$, $\xi_1 = *****1*$, $\xi_2 = 00010011110$, $\xi_3 = *100*01001*$. Definimos el orden de un esquema N_2 como el número de símbolos diferentes de * en el esquema. Para los ejemplos anteriores tenemos $N_2(\xi_0) = 7$, $N_2(\xi_1) = 1$, $N_2(\xi_2) = 11$ y $N_2(\xi_3) = 8$. A un esquema de orden k lo llamaremos un k -esquema. Otra medida importante para el estudio de los esquemas es la longitud definitoria que es la distancia (el número de posiciones) entre el primer y el último bits definidos del esquema. Nuevamente para los ejemplos anteriores $l(\xi_0) = 9$, $l(\xi_1) = 0$, $l(\xi_2) = 10$ y $l(\xi_3) = 8$. Observemos que un esquema de orden N_2 “representa” 2^{N-N_2} cadenas de longitud N , por ejemplo el esquema $*1*$ para el caso de 3 bits representa a las cadenas 010, 011, 110 y 111.

A pesar de que los esquemas parecen ser una extensión natural del espacio de búsqueda, en el caso binario, el espacio de los esquemas tiene 3^N elementos mientras que la base de los bloques constructores (es decir esquemas que solo tienen asteriscos en el extremo izquierdo o alternativamente en el derecho) solo tiene 2^N , así que la base de esquemas es sobrecompleta.

Al igual que a las cadenas (fig. 2.4), podemos darle una representación geométrica a los esquemas como se muestra en la figura 4.2. El espacio de búsqueda es también un hipercubo N -dimensional. Los N -esquemas (cadenas) corresponden a los vértices, los $(N - 1)$ -esquemas son las aristas, los $(N - 2)$ -esquemas corresponden a planos y en general un $N - k$ -esquema corresponde a un $N - k$ -plano, o plano de dimensión $N - k$.

Probablemente el de esquemas es el cambio de escala más usado. Consiste en agrupar a las cadenas que coincidan con ξ en sus bits definidos, siendo ξ un esquema dado. El nivel más definido de un esquema ξ es una cadena, y el menos definido es el esquema con puros asteriscos; este último es el caso trivial que corresponde a toda la población, es por eso que nos detenemos un paso antes y decimos que el nivel menos definido es el de los 1-esquemas.

En el caso de esquemas los cambios de escala no triviales están dados al afectar un esquema de tal modo que disminuya N_2 . El nivel más definido corresponde a $N_2 = N$, es decir, todos los bits definidos. El menos definido corresponde a $N_2 = 1$, a los 1-esquemas. Si consideramos la siguiente sucesión de cambio de escala:

$$C \rightarrow \Gamma_i^{(1)} \rightarrow \Gamma_i^{(2)} \rightarrow \Gamma_i^{(3)} \rightarrow \Gamma_i^{(4)} \dots \quad (4.8)$$

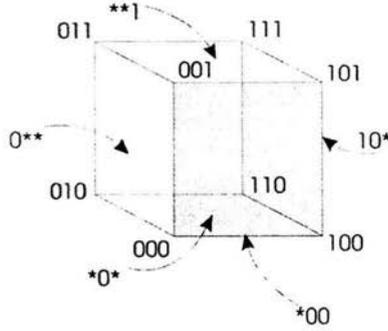


Figura 4.2: Interpretación geométrica de esquemas

donde el superíndice de Γ_i corresponde al índice del cambio de escala. Entonces observamos las siguientes relaciones: $N \geq N_2^{(1)} \geq N_2^{(2)} \geq \dots \geq 1$ y $N \geq l^{(1)} \geq l^{(2)} \geq \dots$. Si denotamos por $\mathcal{R}(N_2, l)$ el conjunto de esquemas de orden N_2 y longitud definitoria l entonces la ecuación (4.8) puede ser reescrita como: $\mathcal{R}(1, 1) = \dots \mathcal{R}(N_2^3, l_3) \mathcal{R}(N_2^2, l_2) \mathcal{R}(N_2^1, l_1)$. A continuación exploraremos el comportamiento de los operadores genéticos más comunes, esta vez dentro del contexto de los esquemas.

4.3.1. Solo selección

Si solo consideramos el efecto de la selección proporcional observamos una covarianza en la forma de la ecuación para selección:

$$P(\xi, t+1) = \sum_{C_i \in \xi} P(C_i, t+1) = \sum_{C_i \in \xi} \frac{f(C_i)}{\bar{f}(t)} P(C_i, t) = \frac{f(\xi, t)}{\bar{f}(t)} P(\xi, t) \quad (4.9)$$

4.3.2. Selección y mutación

Consideremos la ecuación (3.6). Sumamos en ambos lados sobre $C_i \in \xi$.

$$P(\xi, t+1) = \sum_{C_i \in \xi} \mathcal{P}(C_i \rightarrow C_i) P'(C_i, t) + \sum_{C_i \in \xi} \sum_{C_j \neq C_i} \mathcal{P}(C_j \rightarrow C_i) P'(C_j, t) \quad (4.10)$$

Con $\mathcal{P}(C_j \rightarrow C_i)$ definidos como en (3.8). Dividimos la suma sobre C_j en aquellos $C_j \in \xi$ y aquellos que $C_j \notin \xi$

$$\begin{aligned}
P(\xi, t+1) &= \sum_{C_i \in \xi} \left[\prod_{k=1}^N (1-p_m(k)) P'(C_i, t) \right. \\
&+ \sum_{\substack{C_j \neq C_i \\ C_j \in \xi}} \prod_{k \in \{C_j - C_i\}} p_m(k) \prod_{k \in \{C_j - C_i\}_c} (1-p_m(k)) P'(C_j, t) \Big] \\
&+ \sum_{C_i \in \xi} \sum_{\substack{C_j \neq C_i \\ C_j \notin \xi}} \prod_{k \in \{C_j - C_i\}} p_m(k) \prod_{k \in \{C_j - C_i\}_c} (1-p_m(k)) P'(C_j, t)
\end{aligned} \tag{4.11}$$

Ahora dentro de $[\]$ podemos agregar los 2 términos dado que el primer término es simplemente $C_j = C_i$ de la suma. Por lo tanto

$$[\] = \sum_{C_j \in \xi} \prod_{k \in \{C_j - C_i\}} p_m(k) \prod_{k \in \{C_j - C_i\}_c} (1-p_m(k)) P'(C_j, t) \tag{4.12}$$

Ahora, dado que solo estamos considerando los C_j que son lo mismo que C_i en los bits dentro del esquema ξ podemos sacar un factor de $\prod_{\substack{k=1 \\ k \in \xi}}^{N_2} (1-p_m(k))$ dado que $p_m(k)$ solo puede actuar en los bits fuera del esquema.

$$[\] = \prod_{\substack{k=1 \\ k \in \xi}}^{N_2} (1-p_m(k)) \sum_{C_j \in \xi} \prod_{\substack{k \in \{C_j - C_i\} \\ k \notin \xi}} p_m(k) \prod_{\substack{k \in \{C_j - C_i\}_c \\ k \notin \xi}} (1-p_m(k)) P'(C_j, t) \tag{4.13}$$

Ahora consideremos $\sum_{C_i \in \xi} [\]$

$$= \prod_{\substack{k=1 \\ k \in \xi}}^{N_2} (1-p_m(k)) \sum_{C_j \in \xi} P'(C_j, t) \left(\sum_{C_i \in \xi} \prod_{\substack{k \in \{C_j - C_i\} \\ k \notin \xi}} p_m(k) \prod_{\substack{k \in \{C_j - C_i\}_c \\ k \notin \xi}} (1-p_m(k)) \right) \tag{4.14}$$

Ahora, sea n la distancia de Hamming entre C_j y C_i en los bits fuera del esquema. Entonces

$$(\) = \sum_{n=0}^{N-N_2} \binom{N-N_2}{n} p_m^n (1-p_m)^{N-N_2-n} = 1 \tag{4.15}$$

dado que hay un máximo de $N - N_2$ para la distancia de Hamming para los bits fuera del esquema y hay $\binom{N-N_2}{n}$ cadenas C_i de distancia n . Esto es para mutación uniforme. Es claro que el resultado es el mismo para mutación no uniforme pues $(\)$ es una partición de la identidad, es simplemente sumar una probabilidad sobre todas las posibilidades. También $\sum_{C_j \in \xi} P'(C_j, t) = P'(\xi, t)$

$$P(\xi, t+1) = \mathcal{P}(\xi \rightarrow \xi)P'(\xi, t) + \sum_{C_i \in \xi} \sum_{\substack{C_j \neq C_i \\ C_j \neq \xi}} \prod_{k \in \{C_j - C_i\}} p_m(k) \prod_{k \in \{C_j - C_i\}_c} (1 - p_m(k))P'(C_j, t) \quad (4.16)$$

Ahora dividimos los k en aquellos que corresponden a los bits dentro del esquema y aquellos que corresponden a los bits fuera del esquema. Es decir (ξ es un esquema con los mismos bits definidos que ξ y que difiere en estos en al menos un bit).

$$\begin{aligned} & \prod_{k \in \{C_j - C_i\}} p_m(k) \prod_{k \in \{C_j - C_i\}_c} (1 - p_m(k)) = \\ & \prod_{k \in \{\xi - \xi_i\}} p_m(k) \prod_{k \in \{\{C_j - \xi_i\} - \{C_i - \xi_i\}\}} p_m(k) \\ & \times \prod_{k \in \{\xi - \xi_i\}_c} (1 - p_m(k)) \prod_{k \in \{\{C_j - \xi_i\} - \{C_i - \xi_i\}\}_c} (1 - p_m(k)) \end{aligned} \quad (4.17)$$

Considerando el segundo y cuarto término (de los bits fuera del esquema) $\sum_{C_i \in \xi}$ es una suma sobre todas las posibilidades de los bits fuera del esquema y por lo tanto una vez más una partición de la identidad. Por lo tanto

$$P(\xi, t+1) = \mathcal{P}(\xi \rightarrow \xi)P'(\xi, t) + \sum_{\substack{C_j \neq C_i \\ C_j \neq \xi}} \prod_{k \in \{\xi - \xi_i\}} p_m(k) \prod_{k \in \{\xi - \xi_i\}_c} (1 - p_m(k))P'(C_j, t) \quad (4.18)$$

Ahora, $\sum_{\substack{C_j \neq C_i \\ C_j \neq \xi}} p_m(k)$ puede ser escrita como $\sum_{\substack{\xi_i \neq \xi \\ \text{bits dentro} \\ \text{del esquema}}} \sum_{\substack{C_j \in \xi \\ \text{bits fuera} \\ \text{del esquema}}} y$
 $\sum_{C_j \in \xi_i} P'(C_j, t) = P'(\xi_i, t)$ con lo que obtenemos:

$$P(\xi, t+1) = \sum_{C_i \in \xi} P(C_i, t+1) = \mathcal{P}(\xi \rightarrow \xi)P'(\xi, t) + \sum_{\xi_i} \mathcal{P}(\xi_i \rightarrow \xi)P'(\xi_i, t) \quad (4.19)$$

donde ξ_i es cualquier esquema que difiera en al menos 1 bit definitorio de ξ , y la probabilidad $\mathcal{P}(\xi_i \rightarrow \xi)$ de que el esquema ξ_i sea mutado a ξ está dada por

$$\mathcal{P}(\xi_i \rightarrow \xi) = \prod_{k \in \{\xi_i - \xi\}} p_m(k) \prod_{k \in \{\xi_i - \xi\}_c} (1 - p_m(k)) \quad (4.20)$$

Así, hemos obtenido una ecuación muy similar en forma a la de evolución de cadenas (3.6). Esto nos dice que el cambio de escala de esquemas es una manera natural, en el sentido antes mencionado, de extender las ecuaciones para cadenas.

4.3.3. Selección y cruza

A partir de la ecuación (3.29) a continuación derivamos la equivalente para esquemas. Denotemos:

$$\sum_{C_i \in \xi} P(C_i, t+1) = P(\xi_i, t+1) \quad (4.21)$$

Al lado derecho tenemos

$$\sum_{C_i \in \xi} P'(C_i, t)(1-p_c) = (1-p_c)P'(\xi_i, t) = (1-p_c \frac{(l-1)}{N-1} + p_c \frac{(l-1)}{N-1} - p_c)P'(\xi_i, t) \quad (4.22)$$

y

$$\begin{aligned} & \sum_{C_i \in \xi} \frac{p_c}{N-1} \sum_{k=1}^{N-1} P'(C_i^L(k), t)P'(C_i^R(k), t) \\ &= \frac{p_c}{N-1} \sum_{C_i \in \xi} \left(\sum_{k=1}^{l-1} + \sum_{\substack{\text{puntos de cruza} \\ \text{fuera de } \xi}} \right) P'(C_i^L(k), t)P'(C_i^R(k), t) \end{aligned} \quad (4.23)$$

Sumando los dos términos obtenemos:

$$\begin{aligned} & (1-p_c \frac{(l-1)}{N-1})P'(\xi_i, t) - \frac{p_c}{N-1} \sum_{\substack{\text{puntos} \\ \text{fuera de } \xi}} P'(\xi_i, t) \\ & + \sum_{C_i \in \xi} \sum_{k \in \xi_c} \frac{p_c}{N-1} P'(C_i^L(k), t)P'(C_i^R(k), t) + \\ & \frac{p_c}{N-1} \sum_{C_i \in \xi} \sum_{k'=1}^{l-1} P'(C_i^L(k), t)P'(C_i^R(k), t) \end{aligned} \quad (4.24)$$

Ahora, en el tercer término estamos cortando fuera del esquema entonces todo el esquema está en L o R. Entonces estamos sumando sobre todas las cadenas que contienen ξ que está en L o R pero esto significa que en la otra mitad (la que no contiene a ξ) estamos sumando sobre todas las posibilidades. Miremos nuevamente el último término, tenemos:

$$\frac{p_c}{N-1} \sum_{k'=1}^{l-1} \sum_{C_i \in \xi} \sum_{C_j \in C_i^L} \sum_{C_l \in C_i^R} P'(C_j, t)P'(C_l, t) = \frac{p_c}{N-1} \sum_{k'=1} P'(\xi^L, t)P'(\xi^R, t) \quad (4.25)$$

Lo podemos ver de la siguiente manera:

$$\sum_{C_i \in \xi} \sum_{C_j \in C_i^L} \sum_{C_l \in C_i^R} = \sum_{C_i} \prod_{n_1 \in \xi}^{l-1} \delta_{C_i(n_1)\xi(n_1)} \sum_{C_j} \prod_{n_2 \in C_i^L} \delta_{C_j(n_2)C_i^L(n_2)} \sum_{C_l} \prod_{n_3 \in C_i^R} \delta_{C_l(n_3)C_i^R(n_3)} \quad (4.26)$$

Ahora

$$\prod_{n_1 \in \xi}^{l-1} \delta_{C_i(n_1)\xi_L(n_1)} = \prod_{n_1 \in \xi_L} \delta_{C_i(n_1)\xi_L(n_1)} \prod_{n'_1 \in \xi_R} \delta_{C_i(n'_1)\xi_R(n'_1)} \quad (4.27)$$

El término $\prod_{n_2 \in C_i^L} \delta_{C_j(n_2)C_i^L(n_2)} \prod_{n_1 \in \xi_L} \delta_{C_i(n_1)\xi_L(n_1)}$ restringe la suma sobre los C_j 's a aquellos que contienen ξ_L ; similarmente $\prod_{n'_1 \in \xi_R} \delta_{C_i(n'_1)\xi_R(n'_1)} \prod_{n_3 \in C_i^R} \delta_{C_i(n_3)C_i^R(n_3)}$ restringe la suma sobre los C_i 's a aquellos que contienen ξ_R .

Ahora \sum_{C_i} , la suma es sin restricciones, contiene 2^N términos, por lo tanto tenemos:

$$\sum_{C_i} = \sum_{C_i^L} \sum_{C_i^R} \quad (4.28)$$

por lo tanto

$$\sum_{C_i^L} \sum_{C_i^R} \prod_{n_1 \in \xi_L} \delta_{C_i(n_1)\xi_L(n_1)} \prod_{n'_1 \in \xi_R} \delta_{C_i(n'_1)\xi_R(n'_1)} P'(C_i^L(k), t) P'(C_i^R(k), t) = \sum_{C_i \in \xi} P'(C_i^L(k), t) P'(C_i^R(k), t) \quad (4.29)$$

Ahora:

$$\begin{aligned} \sum_{C_i^L} \prod_{n_1 \in \xi_L} \delta_{C_i(n_1)\xi_L(n_1)} P'(C_i^L, t) &= P'(\xi_L, t) \\ \sum_{C_i^R} \prod_{n'_1 \in \xi_R} \delta_{C_i(n'_1)\xi_R(n'_1)} P'(C_i^R, t) &= P'(\xi_R, t) \end{aligned} \quad (4.30)$$

por lo tanto

$$\frac{p_c}{N-1} \sum_{C_i \in \xi} \sum_{k'=1}^{l-1} P'(C_i^L, t) P'(C_i^R, t) = \frac{p_c}{N-1} \sum_{k'=1}^{l-1} P'(\xi_L, t) P'(\xi_R, t) \quad (4.31)$$

Para el segundo y tercer términos en (4.24), consideremos el tercer.

$$\sum_{C_i \in \xi} P'(C_i^L, t) P'(C_i^R, t) = \sum_{C_i^L} \sum_{C_i^R} \prod_{n \in \xi} \delta_{C_i(n)\xi(n)} P'(C_i^L, t) P'(C_i^R, t) \quad (4.32)$$

Ahora, en este término estamos considerando un punto de cruce que corta fuera del esquema, si corta a la derecha entonces todo ξ está en C_i por lo tanto la suma sobre C_i^R es simplemente 1. Es decir $\sum_{C_i^R} P'(C_i^R, t) = 1$ dado que estamos sumando sobre todas las posibles C_i^R sin restricción. Entonces (4.32) nos da $P'(\xi, t)$ y por lo tanto se cancela con el segundo término. Entonces hemos derivado la ecuación de evolución para selección y cruce en términos de esquemas:

$$P(\xi, t+1) = \left(1 - p_c \frac{l-1}{N-1}\right) P'(\xi, t) + \frac{p_c}{N-1} \sum_{k=1}^{l-1} P'(\xi_L(k), t) P'(\xi_R(k), t) \quad (4.33)$$

En términos de máscaras, si sumamos sobre $C_i \in \xi$ en ambos lados de la ecuación (3.29) tenemos

$$P_c(\xi, t) = \left(1 - p_c \frac{\sum_{m \in \mathcal{M}_r(l, N_2)} p_c(m)}{N_{\mathcal{M}}} \right) P'(\xi, t) + \sum_{m \in \mathcal{M}_r(l, N_2)} p_c(m) P'(\xi^s(m), t) P'(\xi^c(m), t) \quad (4.34)$$

donde ξ^c representa la parte del esquema ξ heredada del primer padre y ξ^s es la parte heredada del segundo. $N_{\mathcal{M}_r}(l, N_2)$ es el número de máscaras que afectan a ξ , siendo \mathcal{M}_r el conjunto de dichas máscaras. $N_{\mathcal{M}}$ es el número total de máscaras con $p_c(m) \neq 0$. Notemos que el término $\frac{\sum_{m \in \mathcal{M}_r(l, N_2)} p_c(m)}{N_{\mathcal{M}}}$ renormaliza la probabilidad de cruce manteniendo la covarianza de forma con respecto a la ecuación de cadenas. Nuevamente, la ecuación resultante de la transformación tiene una forma que nos indica que los esquemas son una buena extensión, que de hecho nos da una mejor visión con ecuaciones más elegantes. Mediante dicha extensión podemos estudiar de manera intuitiva la evolución de grupos de cadenas. La ecuación nos sugiere que una buena solución se forma a partir de dos bloques más fundamentales (o menos definidos), uno a la izquierda y otro a la derecha (o en general, uno de s y otro de c) que a su vez fueron creados por otros bloques menos definidos y así sucesivamente.

4.3.4. Ecuación general (selección, mutación y cruce)

Como antes, la ecuación general es el resultado de integrar las ecuaciones para selección mutación y cruce que obtuvimos de manera independiente. Es decir

$$P(\xi, t+1) = \mathcal{P}(\xi \rightarrow \xi) P(\xi, t) + \sum_{\xi'} \mathcal{P}(\xi \rightarrow \xi') P_c(\xi', t) \quad (4.35)$$

donde

$$P_c(\xi, t) = \left(1 - p_c \frac{N_{\mathcal{M}_r}(l, N_2)}{N_{\mathcal{M}}} \right) P'(\xi, t) + \sum_{m \in \mathcal{M}_r(l, N_2)} p_c(m) P'(\xi^s(m), t) P'(\xi^c(m), t) \quad (4.36)$$

De esta sección concluimos que existe una dualidad entre las ecuaciones para cadenas y las que actúan sobre esquemas. Esto nos da la pauta para inferir que los resultados obtenidos en las ecuaciones para esquemas serán válidos también para cadenas (salvo algún término de normalización) y viceversa.

4.4. Mapeo genotipo-fenotipo

Un caso particular de cambio de escala es agrupar las cuerdas con aptitud común, esta agrupación nos permite aprovechar la simetría de la aptitud como

posible simplificación de las ecuaciones. La simetría se observa sin duda en la selección, pero en esta sección estudiaremos la forma en que los otros operadores genéticos comunes rompen dicha simetría o bien la preservan.

4.4.1. Escenario plano

Dado que todas las cuerdas tienen asociado el mismo valor para la función de aptitud, solo podemos hablar de un fenotipo. Es por eso que consideramos a este caso como trivial, dado que observamos el máximo grado de degeneración en la función de aptitud.

4.4.2. Contando unos

En este caso, si consideramos N bits hay $N + 1$ fenotipos correspondientes al número de unos en la cuerda. Si solo consideramos selección proporcional tenemos:

$$P(C_i, t + 1) = \frac{w^H(C_i)}{\sum_j w^H(C_j)P(C_j)} P(C_i, t) \quad (4.37)$$

Donde w^H denota el peso de Hamming, es decir la distancia de Hamming a la cuerda $\mathbf{0} = 000 \dots 0$. Ahora, sumando sobre las cuerdas de la misma clase, es decir las que tienen el mismo peso de Hamming. Denotemos por W_k al elemento de la cubierta que representa a todas las cuerdas con k 1's, donde $k \in [0, N]$.

$$\sum_{C_i \in W_k} P(C_i, t + 1) = P(W_k, t + 1) = \frac{kP(W_k, t)}{\sum_{l=1}^N lP(W_l, t)} \quad (4.38)$$

Al aplicar este cambio de escala al operador de selección encontramos una covarianza de forma nuevamente. A continuación consideramos el operador de mutación junto con selección. Inicialmente reescribimos la ecuación (3.6) como:

$$P(C_i, t + 1) = \sum_{k=1}^N \frac{k}{\bar{f}(t)} \left(\sum_{C_j \in W_k} (1 - p_m)^{N-d(i,j)} p_m^{d(i,j)} P(C_j, t) \right) \quad (4.39)$$

Como en el caso anterior, podemos hacer la suma sobre las cadenas con el mismo fenotipo, buscando con esta agrupación una ecuación más sencilla en términos de grados de libertad efectivos.

$$P(W_i, t + 1) = \sum_{k=1}^N \frac{k}{\bar{f}(t)} \mathcal{P}_1(W_k \rightarrow W_i) P(W_k, t) \quad (4.40)$$

donde

$$\mathcal{P}_1(W_k \rightarrow W_i) = \sum_{C_i \in W_i} p_m^{d(C_i, C_j)} (1 - p_m)^{N-d(C_i, C_j)} \quad (4.41)$$

con $C_j \in W_k$. Notemos que la última ecuación está simplificada de modo que solo se considera una vez a C_j ; esto es por que para cualesquiera $C_i, C_i' \in$

W_k , si existe $C_s \in W_m$ tal que $d^H(C_i, C_s) = r$, entonces existe $C_{s'} \in W_m$ tal que $d^H(C_{i'}, C_{s'}) = r$. Esto nos permite factorizar el término \mathcal{P}_1 . Esta es una propiedad de la distancia de Hamming que se observó en el proceso. La derivación se muestra en la sección A.6. Una ecuación equivalente se muestra en [21], donde Jon Rowe hace un análisis de las funciones que solo dependen del número de unos de la cuerda.

Desafortunadamente para nuestros fines, a pesar de la simplificación antes mencionada no podemos hablar de una invariancia de forma. El operador de mutación actúa rompiendo la simetría inducida por el cambio de escala en el operador de selección.

Por ejemplo para 2 bits tenemos:

$$\begin{aligned} \begin{pmatrix} P(W_0, t+1) \\ P(W_1, t+1) \\ P(W_2, t+1) \end{pmatrix} &= \frac{1}{\bar{f}(t)} \begin{pmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 2 \end{pmatrix} \\ \times \begin{pmatrix} (1-p_m)^2 & 2p_m(1-p_m) & p_m^2 \\ p_m(1-p_m) & (1-p_m)^2 + p_m^2 & p_m(1-p_m) \\ p_m^2 & 2p_m(1-p_m) & (1-p_m)^2 \end{pmatrix} &\begin{pmatrix} P(W_0, t) \\ P(W_1, t) \\ P(W_2, t) \end{pmatrix} \end{aligned} \quad (4.42)$$

En el caso de contando 1's la ecuación para evolución de clases de equivalencia adquiere la siguiente forma:

$$P(W_r, t+1) = \sum_{k=1}^{N-1} \sum_{i=0}^r \eta(k, i) p_c P'(W_i^{L(k)}, t) P'(W_{r-i}^{R(k)}, t) \quad (4.43)$$

Donde $W_i^{L(k)}$ (alternativamente $W_i^{R(k)}$) es la clase de equivalencia de las cadenas que tienen i unos a la izquierda (alternativamente a la derecha) del punto de cruce k .

$$\eta(k, i) = \begin{cases} 1 & \text{si } i \leq k \text{ y } r-i \leq n-k \\ 0 & \text{de otra manera} \end{cases}$$

4.4.3. Modelo de Eigen

Para este modelo solo tenemos dos posibles valores para el fenotipo: el que corresponde a la secuencia maestra y el de toda la demás población. Es decir, al considerar el mapeo genotipo-fenotipo, reducimos el espacio de estados de 2^N posibles a solo 2 grados de libertad efectivos. Consideremos nuevamente solo selección, tenemos:

$$P(C_i, t+1) = \frac{f(C_i)}{\sum_{j \neq sm} f(C_j) P(C_j) + f(C_{sm}) P(C_{sm})} P(C_i, t) \quad (4.44)$$

Este tipo de cambio de escala con selección tiene un alto grado de simetría que puede ser roto por otros operadores como mutación y cruce. Para simplificar la notación denotemos por C_{sm} a la cuerda que denominamos la secuencia maestra,

S_0 al conjunto de cadenas que no son la secuencia maestra. Denotemos $f(C_{sm}) = f_{sm}$ y $f_0 = f(C_j)$ con $j \neq sm$, si sumamos sobre las cuerdas de la población diferentes de la secuencia maestra tenemos:

$$\sum_{C_i \in S_0} P(C_i, t+1) = P(S_0, t+1) = \frac{f(S_0, t)}{f_0 P(S_0, t) + f_{sm} P(C_{sm}, t)} P(S_0, t) \quad (4.45)$$

Ahora introducimos los elementos para considerar la mutación de cuerdas, reescribimos la ecuación (3.6):

$$\begin{aligned} P(C_{sm}, t+1) &= (1-p_m)^N \frac{f_{sm}}{\bar{f}(t)} P(C_{sm}, t) \\ &+ \frac{f_0}{\bar{f}(t)} \sum_{C_j \neq C_{sm}} (1-p_m)^{N-d(C_{sm}, C_j)} p_m^{d(C_{sm}, C_j)} P(C_j, t) \end{aligned} \quad (4.46)$$

$$\begin{aligned} P(C_i, t+1) &= (1-p_m)^N \frac{f_0}{\bar{f}(t)} P(C_i, t) \\ &+ \frac{f_0}{\bar{f}(t)} \sum_{C_j \neq C_{sm}} (1-p_m)^{N-d(C_i, C_j)} p_m^{d(C_i, C_j)} P(C_j, t) \\ &+ \frac{f_{sm}}{\bar{f}(t)} (1-p_m)^{N-d(C_{sm}, C_i)} p_m^{d(C_{sm}, C_i)} P(C_{sm}, t) \end{aligned} \quad (4.47)$$

Si sumamos sobre todas las cuerdas diferentes de la secuencia maestra en esta última ecuación tendremos:

$$\begin{aligned} \sum_{C_i \in S_0} P(C_i, t+1) &= P(S_0, t+1) = (1-p_m)^N \frac{f_0}{\bar{f}(t)} P(S_0, t) \\ &+ \frac{f_0}{\bar{f}(t)} \sum_{C_j \in S_0} \sum_{C_k \in S_0} (1-p_m)^{N-d(C_k, C_j)} p_m^{d(C_k, C_j)} P(C_k, t) \\ &+ \frac{f_{sm}}{\bar{f}(t)} \sum_{C_j \in S_0} (1-p_m)^{N-d(C_{sm}, C_j)} p_m^{d(C_{sm}, C_j)} P(S_m, t) \end{aligned} \quad (4.48)$$

Notemos que este cambio de escala no es natural en el sentido de la covarianza de forma. Consideramos entonces el cambio de escala siguiente: agruparemos todas las cuerdas que tienen la misma distancia de Hamming a la secuencia maestra. Denotemos por D_k al conjunto de todas las cadenas en la población tales que su distancia a la secuencia maestra es k . Esto es $D_k = \{C_j | d^H(C_j, C_{sm}) = k\}$, y podemos reescribir la ecuación para la secuencia maestra considerando mutación

$$P(C_{sm}, t+1) = \frac{f_{sm}}{\bar{f}(t)} P(C_{sm}, t) + \frac{f_0}{\bar{f}(t)} \sum_{k=1}^N (1-p_m)^{N-k} p_m^k P(D_k, t) \quad (4.49)$$

Observamos un cambio de escala más natural en el caso de la secuencia maestra, pero no así para las demás cuerdas. Las ecuaciones anteriores tienen como

objetivo encontrar simplificaciones a las ecuaciones, cosa que en algunos casos se logra. Sin embargo nos hemos dado cuenta que es difícil obtener un cambio de escala que mantenga las propiedades del usual.

El mapeo genotipo-fenotipo puede dar lugar a importantes simplificaciones en las ecuaciones de evolución aprovechando la simetría que induce sobre el operador de selección; sin embargo, la aplicación de otros operadores rompe esta simetría. Esto disminuye las probabilidades de encontrar una forma elegante para las ecuaciones al menos en forma exacta. De lo anterior podemos concluir que, si bien el cambio de escala basado en esquemas ha probado ser una forma elegante de expresar las ecuaciones de evolución, otras alternativas están disponibles para casos particulares. Los cambios de escala tienen como objetivo encontrar los grados de libertad efectivos del problema. En el caso del mapeo genotipo-fenotipo surgen de manera natural de la información del escenario de aptitud, aprovechando la simetría para simplificar las ecuaciones. En el caso de los esquemas, como hemos visto, da lugar a importantes simplificaciones en las ecuaciones que involucran al operador de cruce.

Capítulo 5

Solución de las ecuaciones para pocos grados de libertad en algunos escenarios típicos

En este capítulo replanteamos las ecuaciones derivadas anteriormente, esta vez restringiendo nuestra atención a 1, 2 y 3 bits en los siguientes escenarios representativos: contando unos, modelo de Eigen y escenario plano. Esto con el fin de hacer explícita la complejidad de las soluciones aun para estos casos sencillos. Además, pocos grados de libertad nos permiten desarrollar una mejor intuición del comportamiento del sistema. Es importante notar que, el que los modelos sean de pocos bits no implica que los problemas a tratar deban tener esta cantidad de grados de libertad, lo que pretendemos es dar una visión local del problema, es decir, si trabajamos con cadenas grandes, en un cierto grupo de bits, la función de aptitud puede comportarse linealmente, en otro grupo de manera “needle-in-a-haystack”, etcétera.

En cada una de las siguientes secciones presentamos inicialmente las ecuaciones correspondientes a la dinámica de los operadores y en las subsecciones correspondientes particularizamos a cada uno de los escenarios antes mencionados. Consideramos selección proporcional, mutación con una distribución de probabilidad uniforme y cruce de un punto. Y por último, mostramos la solución correspondiente haciendo énfasis en el movimiento de la población en el simplex para casos límite.

5.1. Soluciones para selección 1 bit

En el caso de un bit el operador de selección es ciertamente trivial y lo incluimos solo como punto de partida. Las siguientes ecuaciones muestran la

evolución de las cadenas:

$$P(i, t + 1) = \frac{f(i)}{j(i)} P(i, t) \quad (5.1)$$

donde $i \in \{0, 1\}$ es una cadena de 1 bit.

5.1.1. Escenario plano

En este caso simplemente no hay manera de generar cambios en la población, es decir, el escenario es completamente estático. Por lo tanto la solución es la siguiente:

$$P(0, t) = P(0, 0) \quad (5.2)$$

$$P(1, t) = P(1, 0) \quad (5.3)$$

Es decir, el único punto fijo es precisamente la condición inicial. Dado que esta condición es arbitraria en principio, el conjunto de posibles puntos fijos es denso en el simplex. El resultado cambia si consideramos una población finita, pues en ese caso tarde o temprano una de las 2 cadenas domina en proporción.

5.1.2. Escenario de contando 1's

$$P(0, t) = 0 \quad (5.4)$$

$$P(1, t) = 1 \quad (5.5)$$

para $t > 0$. En este caso notamos que se alcanza el punto de equilibrio en un solo paso dado que la probabilidad de seleccionar 0 es nula. Este punto fijo es estable y es el único del sistema. Para estudiar el límite cuando la presión selectiva es muy pequeña, consideramos que la aptitud de la cadena 0 es a y el de 1 es $1 + a$. Si iteramos, las soluciones son:

$$P(0, t) = \frac{a^t P(0,0)}{a^t P(0,0) + (1+a)^t P(1,0)} \quad (5.6)$$

$$P(1, t) = \frac{(1+a)^t P(1,0)}{a^t P(0,0) + (1+a)^t P(1,0)} \quad (5.7)$$

Como asumimos a pequeño, la primera ecuación decae mientras la otra predomina con la siguiente razón $\frac{P(0,t)}{P(1,t)} = \left(\frac{a}{1+a}\right)^t \frac{P(0,0)}{P(1,0)}$, aproximándose rápidamente al punto fijo. La velocidad está determinada por el valor de a , entre menor sea, más rápido se llega al punto fijo partiendo del cociente de las proporciones iniciales.

Particularizando aun más, podemos considerar una distribución inicial aleatoria y las soluciones son

$$P(0, t) = \frac{a^t}{a^t + (1+a)^t} \quad (5.8)$$

$$P(1, t) = \frac{(1+a)^t}{a^t + (1+a)^t} \quad (5.9)$$

Así podemos observar el comportamiento de las soluciones en la figura 5.1. Notemos que a sirve como un parámetro de desaceleración de la convergencia, entre mayor sea este parámetro, más suave será la curva; por lo contrario, si $a \rightarrow 0$ habrá un cambio abrupto en la primera generación. Si lo trasladamos al simplex, observamos que la tendencia es ir al vértice con más aptitud.

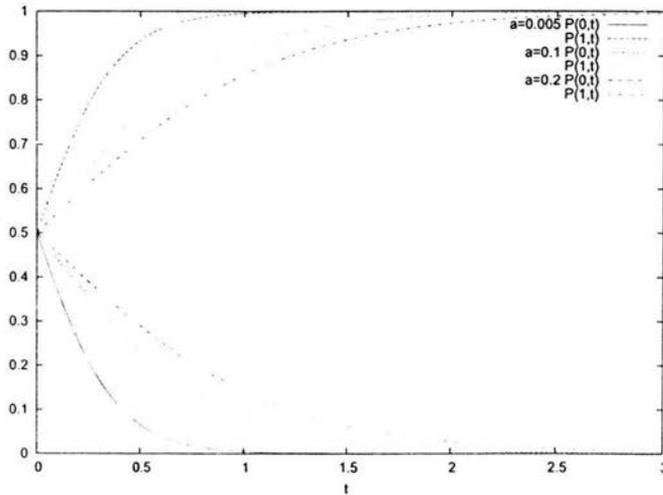


Figura 5.1: Selección para 1 bit contando unos: Comportamiento asintótico de $P(0, t)$ y $P(1, t)$

5.1.3. Modelo de Eigen

Sin importar cual de las cadenas tomemos como secuencia maestra, dado que solo hay otra opción, no podemos apreciar las características que distinguen al modelo de Eigen y las ecuaciones se parecen a un caso menos extremo de contando unos y el punto de equilibrio se puede encontrar en unas pocas iteraciones.

$$P(0, t + 1) = \frac{f_u}{f(t)} P(0, t) \quad (5.10)$$

$$P(1, t + 1) = \frac{f_m}{f(t)} P(1, t) \quad (5.11)$$

Si sustituimos $P(0, t)$ por $1 - P(1, t)$ tenemos

$$P(0, t+1) = \frac{f_0}{f_0 + P(1, t)(f_{sm} - f_0)} P(0, t) \quad (5.12)$$

$$P(1, t+1) = \frac{f_{sm}}{f_0 + P(1, t)(f_{sm} - f_0)} P(1, t) \quad (5.13)$$

Así podemos resolver para $P(1, t)$ iterando. Observamos que: $P(1, 1) = \frac{f_{sm} P(1, 0)}{f_0 + P(1, 0)(f_{sm} - f_0)}$ y $P(1, 2) = \frac{f_{sm}^2 P(1, 0)}{f_0^2 + (f_{sm} - f_0)(f_{sm} + f_0) P(1, 0)}$. Si continuamos iterando observamos que el comportamiento del sistema es el siguiente:

$$P(0, t) = \frac{f_0^t + f_0 P(1, 0)(f_{sm} - f_0) \left(\sum_{i=0}^t f_{sm}^i f_0^{t-i} - f_{sm}^t \right)}{f_0^t + P(1, 0)(f_{sm} - f_0) \left(\sum_{i=0}^t f_{sm}^i f_0^{t-i} \right)} = \frac{f_0^t P(0, 0)}{f_0^t P(0, 0) + f_{sm}^t P(1, 0)}$$

$$P(1, t) = \frac{f_{sm}^t P(1, 0)}{f_0^t + P(1, 0)(f_{sm} - f_0) \left(\sum_{i=0}^t f_{sm}^i f_0^{t-i} \right)} = \frac{f_{sm}^t P(1, 0)}{f_0^t P(0, 0) + f_{sm}^t P(1, 0)}$$

Notemos que hasta este punto no hemos hecho ninguna suposición sobre los valores de f_0 y f_{sm} . Por lo que las ecuaciones anteriores son completamente generales para el caso de 1 bit. La generalización de este resultado se presenta en la sección 6.2.

Si asumimos que $f_0 \ll f_{sm}$, la secuencia maestra predominará con la siguiente razón: $\frac{P(0, t)}{P(1, t)} = \left(\frac{f_0}{f_{sm}} \right)^t \frac{P(0, 0)}{P(1, 0)}$. La velocidad con la que se acerquen al punto fijo dependerá de las proporciones iniciales y la diferencia entre las aptitudes. Si ahora particularizamos al caso en el que la población inicial es aleatoria, obtenemos

$$P(0, t) = \left(1 + \left(\frac{f_{sm}}{f_0} \right)^t \right)^{-1} \quad (5.14)$$

$$P(1, t) = \left(1 + \left(\frac{f_0}{f_{sm}} \right)^t \right)^{-1} \quad (5.15)$$

Podemos observar los resultados en la figura 5.2. A pesar de que en este caso es más difícil tener un cambio abrupto en el tiempo continuo, para efectos del tiempo discreto, los resultados son prácticamente los mismos, es decir, la convergencia está asegurada con un número muy reducido de iteraciones.

5.2. Soluciones para selección 2 bits

Para dos bits ya notamos más opciones dado que hay más diversidad. Las ecuaciones de la dinámica para solo selección son las siguientes:

$$P(i, j, t+1) = \frac{f(i, j)}{f(t)} P(i, j, t) \quad (5.16)$$

con $i, j \in \{0, 1\}$.

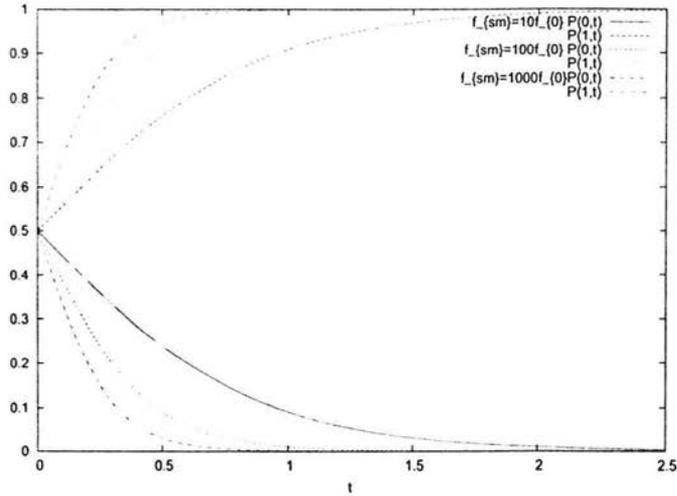


Figura 5.2: Selección para 1 bit modelo de Eigen: Comportamiento asintótico de $P(0, t)$ y $P(1, t)$

5.2.1. Escenario plano

Como observamos antes, el caso del escenario plano no hay más que decir, las proporciones iniciales permanecen estáticas y se convierten por tanto en el punto fijo de las ecuaciones.

$$P(ij, t) = P(ij, 0) \quad (5.17)$$

5.2.2. Escenario de contando 1's

$$P(00, t + 1) = 0 \quad (5.18)$$

$$P(01, t + 1) = \frac{P(01, t)}{\bar{f}(t)} \quad (5.19)$$

$$P(10, t + 1) = \frac{P(10, t)}{\bar{f}(t)} \quad (5.20)$$

$$P(11, t + 1) = \frac{2P(11, t)}{\bar{f}(t)} \quad (5.21)$$

Donde $\bar{f}(t) = P(01, t) + P(10, t) + 2P(11, t)$. En estas ecuaciones notamos que la decadencia del 00 es inmediata, pero la de 01 y 10 es más paulatina, mientras 11 crece al doble de velocidad que las anteriores decrecen.

Si pasamos a una base dada por el mapeo genotipo-fenotipo de contando 1's

(como en la sección 4.4.2), tenemos las siguientes ecuaciones:

$$P(W_0, t + 1) = 0 \quad (5.22)$$

$$P(W_1, t + 1) = \frac{P(W_1, t)}{P(W_1, t) + 2P(W_2, t)} \quad (5.23)$$

$$P(W_2, t + 1) = \frac{2P(W_2, t)}{P(W_1, t) + 2P(W_2, t)} \quad (5.24)$$

Notemos que el problema se redujo a 2 grados de libertad efectivos, entonces si sustituimos $P(W_1, t)$ por $1 - P(W_2, t)$ y simplificamos obtenemos:

$$P(W_0, t + 1) = 0 \quad (5.25)$$

$$P(W_1, t + 1) = \frac{1 - P(W_2, t)}{1 + P(W_2, t)} \quad (5.26)$$

$$P(W_2, t + 1) = \frac{2P(W_2, t)}{1 + P(W_2, t)} \quad (5.27)$$

Entonces $P(W_2, 1) = \frac{2P(W_2, 0)}{1 + P(W_2, 0)}$, $P(W_2, 2) = \frac{4P(W_2, 0)}{1 + 3P(W_2, 0)}$. Haciendo algunas iteraciones más, es fácil ver que:

$$P(W_0, t) = 0 \quad (5.28)$$

$$P(W_1, t) = \frac{P(W_1, 0)}{P(W_1, 0) + 2^t P(W_2, 0)} \quad (5.29)$$

$$P(W_2, t) = \frac{2^t P(W_2, 0)}{P(W_1, 0) + 2^t P(W_2, 0)} \quad (5.30)$$

Regresando a la base original, si suponemos una distribución inicial aleatoria, esperamos que $P(01, t) = P(10, t) = \frac{1}{2}P(W_1, t)$. Esto nos muestra la utilidad del cambio de escala para reducir el problema a los grados de libertad efectivos.

5.2.3. Modelo de Eigen

Supongamos que 11 es la secuencia maestra.

$$P(00, t + 1) = \frac{f_0}{f(t)} P(00, t) \quad (5.31)$$

$$P(01, t + 1) = \frac{f_0}{f(t)} P(01, t) \quad (5.32)$$

$$P(10, t + 1) = \frac{f_0}{f(t)} P(10, t) \quad (5.33)$$

$$P(11, t + 1) = \frac{f_{sm}}{f(t)} P(11, t) \quad (5.34)$$

Donde $\bar{f}(t) = f_0(P(00, t) + P(10, t) + P(01, t)) + f_{sm}P(11, t)$. Dada la suposición de que $f_0 \ll f_{sm}$ notemos que de existir la cuerda correspondiente a la secuencia maestra en la población actual, ésta predominará inmediatamente para la siguiente generación a menos que la suma de las proporciones de las otras cadenas sea mucho mayor que la proporción de la secuencia maestra. Al igual que en el escenario de contando 1's podemos hacer un cambio de escala del estilo mapeo genotipo-fenotipo (sección 4.4.3). Tenemos las siguientes ecuaciones:

$$P(S_0, t + 1) = \frac{f_0 P(S_0, t)}{f_0 P(S_0, t) + f_{sm} P(C_{sm}, t)} \quad (5.35)$$

$$P(C_{sm}, t + 1) = \frac{f_{sm} P(C_{sm}, t)}{f_0 P(S_0, t) + f_{sm} P(C_{sm}, t)} \quad (5.36)$$

Debemos notar que las ecuaciones tienen la misma forma que las de (5.10) y por lo tanto la solución es:

$$P(S_0, t) = \frac{f_0^t P(S_0, 0)}{f_0^t P(S_0, 0) + f_{sm}^t P(C_{sm}, 0)} \quad (5.37)$$

$$P(C_{sm}, t) = \frac{f_{sm}^t P(C_{sm}, 0)}{f_0^t P(S_0, 0) + f_{sm}^t P(C_{sm}, 0)} \quad (5.38)$$

Por lo tanto, regresando a la base original, debemos tener que $P(00, t) = P(01, t) = P(10, t) = \frac{1}{3}P(S_0, t)$. De existir la secuencia maestra, esta predomina eventualmente. De hecho con la misma velocidad que en el caso de 1 bit, solo que en esta ocasión compete contra la suma de las proporciones en S_0 .

5.3. Soluciones para selección 3 bits

En el presente trabajo solo consideramos soluciones para las ecuaciones con 3 bits o menos, pues estas son todavía manejables. Sin embargo, pero como es de esperarse, al incrementar la dimensión, las ecuaciones se hacen más complejas e incluso intratables en algunos casos.

$$P(ijk, t + 1) = \frac{f(ijk)}{f(t)} P(ijk, t) \quad (5.39)$$

5.3.1. Escenario plano

Como se mencionó anteriormente, la ausencia de presión selectiva no permite la existencia de una dinámica por lo que el escenario es completamente en estático.

5.3.2. Escenario de contando 1's

Conforme aumentamos el espacio de búsqueda se hacen más evidentes los beneficios del cambio de escala. En este caso, podemos escribir directamente las soluciones en la base de contando 1's:

$$P(W_0, t) = 0 \quad (5.40)$$

$$P(W_1, t) = \frac{P(W_1, 0)}{f(t)} \quad (5.41)$$

$$P(W_2, t) = \frac{2^t P(W_2, 0)}{f(t)} \quad (5.42)$$

$$P(W_3, t) = \frac{3^t P(W_3, 0)}{f(t)} \quad (5.43)$$

donde $\bar{f}(t) = P(W_1, 0) + 2^t P(W_2, 0) + 3^t P(W_3, 0)$. Notemos que mientras aumentamos la longitud de las cadenas la presión selectiva disminuye, es decir, la diferencia entre k y $k + 1$ unos (para k suficientemente grande) es menor relativamente y cada vez más iteraciones son requeridas para que las cadenas con k unos dominen en proporción.

5.3.3. Modelo de Eigen

Supongamos que la secuencia maestra es 111, tenemos:

$$P(ijk, t + 1) = \frac{f_0 P(ijk, t)}{f(t)} \quad (5.44)$$

$$P(111, t + 1) = \frac{f_{sm} P(111, t)}{f(t)} \quad (5.45)$$

Donde $ijk \in \{0, 1\}^3 \setminus \{111\}$ y $\bar{f}(t) = f_0 \sum_{ijk \in \{0, 1\}^3 \setminus \{111\}} P(ijk, t) + f_{sm} P(111, t)$. El caso de 3 bits es tal que no se necesita todavía una diferencia notable entre la aptitud de la cadena de la secuencia maestra y el de las demás para que la secuencia maestra predomine fácilmente. Al igual que antes, las soluciones en la base escalada son las siguientes.

$$P(S_0, t) = \frac{f_0^t P(S_0, 0)}{f_0^t P(S_0, 0) + f_{sm}^t P(C_{sm}, 0)} \quad (5.46)$$

$$P(C_{sm}, t) = \frac{f_{sm}^t P(C_{sm}, 0)}{f_0^t P(S_0, 0) + f_{sm}^t P(C_{sm}, 0)} \quad (5.47)$$

En la figura (5.3) observamos como el cociente de aptitudes ($\frac{f_0}{f_{sm}}$) afecta la convergencia al vértice 111 del simplex, entre menor sea este cociente, la convergencia se acelera mientras que cuando tiende a uno, el tiempo para la convergencia se incrementa dramáticamente.

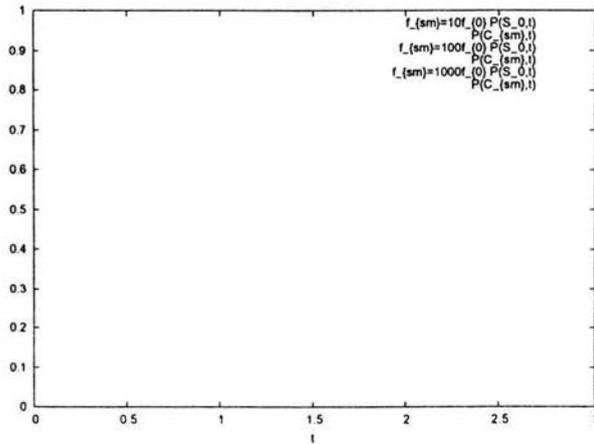


Figura 5.3: Selección para 3 bits modelo de Eigen: Comportamiento asintótico de $P(S_0, t)$ y $P(C_{sm}, t)$

Debemos notar que mientras mayor sea la longitud de las cadenas, es probable que la suma de las proporciones de las cadenas diferentes de la secuencia maestra predomine por más tiempo sobre la proporción de la secuencia maestra, aunque la proporción de la secuencia maestra crecerá exponencialmente con

el tiempo hasta dominar irremediamente (fig. 5.4). Notemos también que el tiempo para la convergencia al vértice 111 del simplex es relativamente corto comparado con el número tan grande de cadenas que se agregan a la búsqueda. De lo anterior podemos concluir que la velocidad de convergencia es función del cociente de las aptitudes $\frac{f_0}{f_{sm}}$, del cociente de proporciones iniciales $\frac{P(C_{sm,0})}{P(S_{0,0})}$ y de la longitud de las cuerdas.

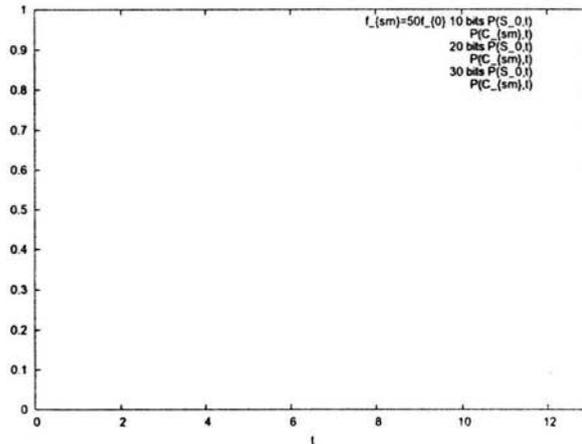


Figura 5.4: Selección para 10, 20 y 30 bits modelo de Eigen: Comportamiento asintótico de $P(S_0, t)$ y $P(C_{sm}, t)$

Notemos que en general dada la diferencia entre la aptitud de la secuencia maestra y la de las demás cuerdas, la predominancia de la secuencia maestra sobre las demás está garantizada en poco tiempo (salvo efectos de población finita). Sin embargo, requerimos de los operadores de exploración para obtener la secuencia maestra, pues de no existir inicialmente en la población, no hay posibilidad de crearla si solo consideramos selección.

5.4. Soluciones para selección y mutación 1 bit

Ahora agreguemos el operador de mutación, de (3.6) tenemos:

$$P(0, t + 1) = P'(0, t)(1 - p_m) + P'(1, t)p_m \quad (5.48)$$

$$P(1, t + 1) = P'(1, t)(1 - p_m) + P'(0, t)p_m \quad (5.49)$$

Podemos expresar estas ecuaciones en términos matriciales de la siguiente forma:

$$\begin{pmatrix} P(0, t + 1) \\ P(1, t + 1) \end{pmatrix} = \begin{pmatrix} 1 - p_m & p_m \\ p_m & 1 - p_m \end{pmatrix} \begin{pmatrix} P'(0, t) \\ P'(1, t) \end{pmatrix} \quad (5.50)$$

5.4.1. Escenario plano

En el caso particular de un escenario plano la ecuación (5.50) se transforma en:

$$\begin{pmatrix} P(1, t) \\ P(0, t) \end{pmatrix} = \begin{pmatrix} 1 - p_m & p_m \\ p_m & 1 - p_m \end{pmatrix}^t \begin{pmatrix} P(1, 0) \\ P(0, 0) \end{pmatrix} \quad (5.51)$$

Y como $\bar{\mathbf{A}}$ es una matriz simétrica, siempre podemos encontrar una matriz ortonormal $\bar{\mathbf{Q}}$ tal que $\bar{\mathbf{Q}}\bar{\mathbf{A}}\bar{\mathbf{Q}} = \text{diag}\{\lambda_1, \dots, \lambda_{2^n}\}$ con $\bar{\mathbf{Q}}$ y λ_i los valores propios, de tal manera que podemos reemplazar a $\bar{\mathbf{A}}$ por una matriz diagonal similar. Podemos usar esto para facilitar los cálculos posteriores de la evolución de las proporciones de las cuerdas.

Lo primero que debemos hacer es calcular el polinomio característico que resulta ser $(\lambda - 1)(\lambda - 1 + 2p_m)$ y los valores propios son: 1 y $1 - 2p$ asociados con los vectores propios $[1/\sqrt{2}, 1/\sqrt{2}]$ y $[1/\sqrt{2}, -1/\sqrt{2}]$ así que podemos construir la matriz de similaridad $\bar{\mathbf{Q}}$ de la siguiente forma:

$$\bar{\mathbf{Q}} = \begin{pmatrix} \frac{\sqrt{2}}{2} & -\frac{\sqrt{2}}{2} \\ \frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} \end{pmatrix} \quad (5.52)$$

Y entonces tenemos:

$$\begin{pmatrix} P(1, t) \\ P(0, t) \end{pmatrix} = \begin{pmatrix} \frac{1}{2}\sqrt{2} & -\frac{1}{2}\sqrt{2} \\ \frac{1}{2}\sqrt{2} & \frac{1}{2}\sqrt{2} \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & (1 - 2p_m)^t \end{pmatrix} \begin{pmatrix} \frac{1}{2}\sqrt{2} & \frac{1}{2}\sqrt{2} \\ -\frac{1}{2}\sqrt{2} & \frac{1}{2}\sqrt{2} \end{pmatrix} \begin{pmatrix} P(1, 0) \\ P(0, 0) \end{pmatrix} \quad (5.53)$$

Y de esta manera hemos encontrado una ecuación iterativa que nos da el número esperado de cadenas a cualquier tiempo. Notemos por otro lado que la matriz $\bar{\mathbf{A}}$ es una matriz de transición para una cadena de Markov regular y por el teorema 2.7.3 sabemos que converge a un punto fijo, por lo tanto tomando el límite cuando t tiende a infinito tenemos la siguiente ecuación:

$$\lim_{t \rightarrow \infty} \begin{pmatrix} P(1, t) \\ P(0, t) \end{pmatrix} = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} \end{pmatrix} \begin{pmatrix} P(1, 0) \\ P(0, 0) \end{pmatrix} = \begin{pmatrix} \frac{1}{2} \\ \frac{1}{2} \end{pmatrix} \quad (5.54)$$

Es decir, las proporciones de las cuerdas terminan por ser las mismas e igual que el promedio. Siendo el centro del simplex el único punto fijo. Notemos también que este resultado es independiente de las condiciones iniciales.

5.4.2. Contando 1's

Si en la ecuación (5.50) asignamos aptitud a a la cadena 0 y $1 + a$ a 1 tenemos:

$$\begin{pmatrix} P(0, t + 1) \\ P(1, t + 1) \end{pmatrix} = \begin{pmatrix} \frac{a(1-p_m)}{aP(0,t)+(1+a)P(1,t)} & \frac{(1+a)p_m}{aP(0,t)+(1+a)P(1,t)} \\ \frac{ap_m}{aP(0,t)+(1+a)P(1,t)} & \frac{(1+a)(1-p_m)}{aP(0,t)+(1+a)P(1,t)} \end{pmatrix} \begin{pmatrix} P(0, t) \\ P(1, t) \end{pmatrix} \quad (5.55)$$

Calculamos sus valores propios $\lambda_1 = (1 - p_m)a + (1 - p_m + \beta_0)/2$ y $\lambda_2 = (1 - p_m)a + (1 - p_m - \beta_0)/2$, con $[\frac{-1+p_m+\beta_0}{2p_m a}, 1]$ y $[\frac{-1+p_m-\beta_0}{2p_m a}, 1]$ sus vectores propios asociados respectivamente, con $\beta_0 = \sqrt{4p_m^2 a^2 + 4p_m^2 a + 1 - 2p_m + p_m^2}$. Con estos vectores como columnas formamos la matriz $\tilde{\mathbf{Q}}$ que cumple la siguiente relación:

$$\mathbf{P}(t+1) = \frac{1}{f(t)} \tilde{\mathbf{Q}} \mathbf{D} \tilde{\mathbf{Q}} \mathbf{P}(t) \quad (5.56)$$

con $\mathbf{P}(t)$ el vector de proporciones y $\mathbf{D} = \text{diag}\{\lambda_1, \lambda_2\}$ y $\tilde{\mathbf{Q}}$ es el inverso multiplicativo de $\tilde{\mathbf{Q}}$, si multiplicamos ambos lados por $\tilde{\mathbf{Q}}$ a la izquierda tenemos:

$$\tilde{\mathbf{Q}} \mathbf{P}(t+1) = \frac{1}{f(t)} \mathbf{D} \tilde{\mathbf{Q}} \mathbf{P}(t) \quad (5.57)$$

Ahora hagamos el siguiente cambio de base: $\alpha(t) = \tilde{\mathbf{Q}} \mathbf{P}$, tenemos entonces que

$$\alpha(t+1) = \frac{1}{f(t)} \mathbf{D} \alpha(t) \quad (5.58)$$

y como \mathbf{D} es diagonal fácilmente podemos iterar el sistema obteniendo:

$$\alpha(t) = \left(\prod_{i=0}^t \frac{1}{f(i)} \right) \mathbf{D}^t \alpha(0) \quad (5.59)$$

De aquí obtenemos 2 ecuaciones, si dividimos $\alpha_2(t)$ entre $\alpha_1(t)$ podemos eliminar la aptitud promedio. Notemos que $\alpha_1(t)$ no se anula en el intervalo $[0, 1]$

$$\frac{\alpha_2(t)}{\alpha_1(t)} = \left(\frac{\lambda_2}{\lambda_1} \right)^t \frac{\alpha_2(0)}{\alpha_1(0)} \quad (5.60)$$

Para resolver el sistema usamos la ecuación de conservación de probabilidad en la nueva base

$$1 = \alpha_2(t) \left(\frac{-1 + p_m + \beta_0}{2p_m a} + 1 \right) + \alpha_1(t) \left(\frac{-1 + p_m - \beta_0}{2p_m a} + 1 \right) \quad (5.61)$$

Así encontramos las soluciones:

$$\alpha_1(t) = \left[\left(\frac{\lambda_2}{\lambda_1} \right)^t \frac{\alpha_2(0)}{\alpha_1(0)} \left(\frac{-1 + p_m + \beta_0}{2p_m a} + 1 \right) + \left(\frac{-1 + p_m - \beta_0}{2p_m a} + 1 \right) \right]^{-1} \quad (5.62)$$

$$\alpha_2(t) = \left(\frac{\lambda_2}{\lambda_1} \right)^t \frac{\alpha_2(0)}{\alpha_1(0)} \alpha_1(t) \quad (5.63)$$

Ahora solo nos resta regresar a la base original mediante la transformación inversa $\mathbf{P} = \tilde{\mathbf{Q}} \alpha$, así obtenemos

$$P(0, t) = \frac{(-1+p_m+\beta_0)\alpha_1+(-1+p_m-\beta_0)\alpha_2}{2p_m a} \quad (5.64)$$

$$P(1, t) = \alpha_1 + \alpha_2 \quad (5.65)$$

Un caso particular es cuando tomamos el límite cuando t tiende a infinito, en este caso podemos ver el comportamiento asintótico de las soluciones. Notemos que el cociente $\frac{\lambda_2(t)}{\lambda_1(t)}$ en valor absoluto es menor que 1 en el intervalo $[0, 1)$ por lo que al tomar potencias grandes tiende a anularse. En las figuras 5.5 y 5.6 se muestran los comportamientos de las cadenas. Notemos que conforme $a \rightarrow 0$ el comportamiento de las soluciones tiende a ser lineal y se converge a un extremo del simplex, mientras que en otros casos se alcanzan diferentes puntos fijos intermedios.

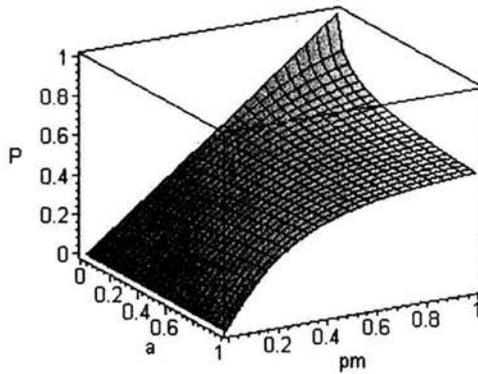


Figura 5.5: Selección y mutación para 1 bit contando unos: Comportamiento asintótico de $P(0, t)$

5.4.3. Modelo de Eigen

$$\begin{pmatrix} P(0, t+1) \\ P(1, t+1) \end{pmatrix} = \begin{pmatrix} 1-p_m & p_m \\ p_m & 1-p_m \end{pmatrix} \begin{pmatrix} P(0, t) \\ P(1, t) \end{pmatrix} \quad (5.66)$$

$$\times \begin{pmatrix} \frac{f_0}{f_0 P(0,t) + f_{sm} P(1,t)} & 0 \\ 0 & \frac{f_{sm}}{f_0 P(0,t) + f_{sm} P(1,t)} \end{pmatrix}$$

Para el modelo de Eigen seguimos el mismo procedimiento que para contando unos. Los valores característicos son: $\lambda_1 = (-p_m f_{sm} + f_0 + f_{sm} - p_m f_0 + \beta_0)/2$ y $\lambda_2 = (-p_m f_{sm} + f_0 + f_{sm} - p_m f_0 - \beta_0)/2$ con

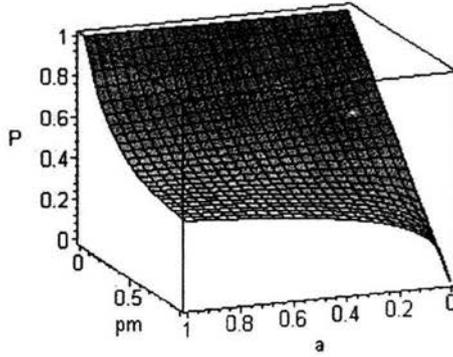


Figura 5.6: Selección y mutación para 1 bit contando unos: Comportamiento asintótico de $P(1, t)$

$\beta_0 = \sqrt{(f_{sm}^2 + f_0^2)(1 - p_m)^2 + 2f_{sm}f_0(-1 + 2p_m + p_m^2)}$ y los vectores característicos son: $\left[\frac{\lambda_1 + f_{sm}(1 - p_m)}{f_0 p_m}, 1\right]$ y $\left[\frac{\lambda_2 + f_{sm}(1 - p_m)}{f_0 p_m}, 1\right]$. El equivalente a la ecuación (5.61) es:

$$1 = \alpha_2(t) \left(\frac{\lambda_1 + f_{sm}(1 - p_m)}{f_0 p_m} + 1 \right) + \alpha_1(t) \left(\frac{\lambda_2 + f_{sm}(1 - p_m)}{f_0 p_m} + 1 \right) \quad (5.67)$$

Y las soluciones al sistema son:

$$\alpha_1(t) = \left[\left(\frac{\lambda_2}{\lambda_1} \right)^t \frac{\alpha_2(0)}{\alpha_1(0)} \left(\frac{\lambda_1 + f_{sm}(1 - p_m)}{p_m f_0} + 1 \right) + \left(\frac{\lambda_2 + f_{sm}(1 - p_m)}{p_m f_0} + 1 \right) \right]^{-1} \quad (5.68)$$

$$\alpha_2(t) = \left(\frac{\lambda_2}{\lambda_1} \right)^t \frac{\alpha_2(0)}{\alpha_1(0)} \alpha_1(t) \quad (5.69)$$

Ahora solo nos resta regresar a la base original mediante la transformación inversa, así obtenemos

$$P(0, t) = \frac{(\lambda_1 + f_{sm}(1 - p_m))\alpha_1 + (\lambda_2 + f_{sm}(1 - p_m))\alpha_2}{p_m f_0} \quad (5.70)$$

$$P(1, t) = \alpha_1 + \alpha_2 \quad (5.71)$$

Como podemos observar, esta solución es la solución general por que no suponemos nada a priori sobre las aptitudes. Ahora consideremos el caso particular del límite cuando t tiende a infinito, así podemos ver el comportamiento asintótico de las soluciones. Igual que antes, notemos que el cociente $\frac{\lambda_2(t)}{\lambda_1(t)}$ en valor absoluto es menor que 1 en el intervalo $[0, 1)$ por lo que al tomar potencias grandes tiende a anularse. En la figura 5.7 se muestran los comportamientos de las cadenas.

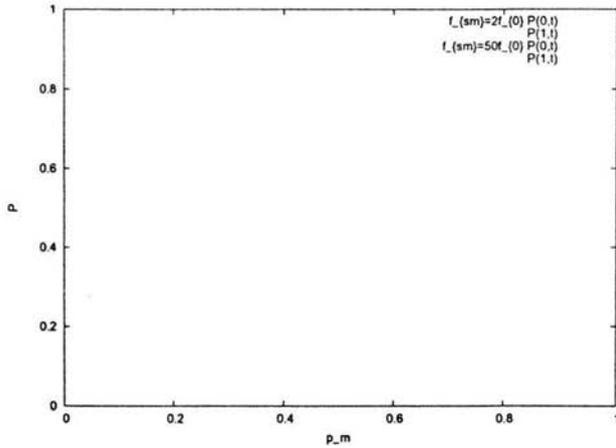


Figura 5.7: Selección y mutación para 1 bit modelo de Eigen: Comportamiento asintótico de $P(0, t)$ y $P(1, t)$. Notemos que cuanto mayor es la diferencia de las aptitudes, más se tiende al comportamiento lineal

5.5. Soluciones para selección y mutación 2 bits

Consideremos el siguiente sistema de ecuaciones en su forma matricial:

$$\begin{pmatrix} P(00, t+1) \\ P(01, t+1) \\ P(10, t+1) \\ P(11, t+1) \end{pmatrix} = \begin{pmatrix} (1-p_m)^2 & (1-p_m)p_m & (1-p_m)p_m & p_m^2 \\ (1-p_m)p_m & (1-p_m)^2 & p_m^2 & (1-p_m)p_m \\ (1-p_m)p_m & p_m^2 & (1-p_m)^2 & (1-p_m)p_m \\ p_m^2 & (1-p_m)p_m & (1-p_m)p_m & (1-p_m)^2 \end{pmatrix} \begin{pmatrix} P'(00, t) \\ P'(01, t) \\ P'(10, t) \\ P'(11, t) \end{pmatrix} \quad (5.72)$$

5.5.1. Escenario plano

Ahora en un Escenario plano de la ecuación (5.72)

$$\begin{pmatrix} P(00, t) \\ P(01, t) \\ P(10, t) \\ P(11, t) \end{pmatrix} = \begin{pmatrix} (1-p_m)^2 & (1-p_m)p_m & (1-p_m)p_m & p_m^2 \\ (1-p_m)p_m & (1-p_m)^2 & p_m^2 & (1-p_m)p_m \\ (1-p_m)p_m & p_m^2 & (1-p_m)^2 & (1-p_m)p_m \\ p_m^2 & (1-p_m)p_m & (1-p_m)p_m & (1-p_m)^2 \end{pmatrix}^t \begin{pmatrix} P(00, 0) \\ P(01, 0) \\ P(10, 0) \\ P(11, 0) \end{pmatrix} \quad (5.73)$$

Y nuevamente, como es una matriz simétrica podemos reemplazarla mediante una transformación de similitud, para facilitar los cálculos. Como antes, obtenemos el polinomio característico que es: $(\lambda - 1)(\lambda - 1 + 2p_m)^2(\lambda - 4p_m^2 + 4p_m - 1)$. Entonces los valores propios son: 1 , $1 - 2p_m$ de multiplicidad 2 y $4p_m^2 - 4p_m + 1$, asociados con los vectores propios $[\frac{1}{2}, \frac{1}{2}, \frac{1}{2}, \frac{1}{2}]$, $[\frac{-1}{\sqrt{2}}, 0, 0, \frac{1}{\sqrt{2}}]$, $[0, \frac{-1}{\sqrt{2}}, \frac{1}{\sqrt{2}}, 0]$ y $[\frac{1}{2}, -\frac{1}{2}, -\frac{1}{2}, \frac{1}{2}]$. Entonces podemos construir a Q de la siguiente manera:

$$\begin{pmatrix} \frac{1}{2} & -\frac{1}{2}\sqrt{2} & 0 & \frac{1}{2} \\ \frac{1}{2} & 0 & -\frac{1}{2}\sqrt{2} & -\frac{1}{2} \\ \frac{1}{2} & 0 & \frac{1}{2}\sqrt{2} & -\frac{1}{2} \\ \frac{1}{2} & \frac{1}{2}\sqrt{2} & 0 & \frac{1}{2} \end{pmatrix} \quad (5.74)$$

Y nuevamente obtenemos una ecuación iterativa para calcular la proporción de cualquier cadena a un tiempo arbitrario basándonos únicamente en los valores iniciales:

$$\begin{pmatrix} P(00, t) \\ P(01, t) \\ P(10, t) \\ P(11, t) \end{pmatrix} = \begin{pmatrix} \frac{1}{2} & -\frac{1}{\sqrt{2}} & 0 & \frac{1}{2} \\ \frac{1}{2} & 0 & -\frac{1}{\sqrt{2}} & -\frac{1}{2} \\ \frac{1}{2} & 0 & \frac{1}{\sqrt{2}} & -\frac{1}{2} \\ \frac{1}{2} & \frac{1}{\sqrt{2}} & 0 & \frac{1}{2} \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & (1-2p_m)^t & 0 & 0 \\ 0 & 0 & (1-2p_m)^t & 0 \\ 0 & 0 & 0 & (1-2p_m)^{2t} \end{pmatrix} \\ \times \begin{pmatrix} \frac{1}{2} & \frac{1}{2} & \frac{1}{2} & \frac{1}{2} \\ -\frac{1}{\sqrt{2}} & 0 & 0 & \frac{1}{\sqrt{2}} \\ 0 & -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0 \\ \frac{1}{2} & -\frac{1}{2} & -\frac{1}{2} & \frac{1}{2} \end{pmatrix} \begin{pmatrix} P(00, 0) \\ P(01, 0) \\ P(10, 0) \\ P(11, 0) \end{pmatrix} \quad (5.75)$$

Nuevamente tomando el límite cuando el tiempo tiende a infinito nos damos cuenta que las proporciones se homogeneizan al promedio (1/4). Este resulta ser un punto fijo estable para estas ecuaciones y es único. En términos de la representación geométrica esto equivale a estacionarse en el punto intermedio $(\frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4})$ alejado de los vértices. En general el punto de equilibrio es $(\frac{1}{2^N}, \dots, \frac{1}{2^N})$

5.5.2. Escenario de contando 1's

Nuevamente, dada una cadena C_i le asignamos el valor de aptitud de $w^H(C_i)$

$$\begin{pmatrix} P(00, t+1) \\ P(01, t+1) \\ P(10, t+1) \\ P(11, t+1) \end{pmatrix} = \frac{1}{\bar{f}(t)} \begin{pmatrix} 0 & (1-p_m)p_m & (1-p_m)p_m & 2p_m^2 \\ 0 & (1-p_m)^2 & p_m^2 & 2(1-p_m)p_m \\ 0 & p_m^2 & (1-p_m)^2 & 2(1-p_m)p_m \\ 0 & (1-p_m)p_m & (1-p_m)p_m & 2(1-p_m)^2 \end{pmatrix} \begin{pmatrix} P(00, t) \\ P(01, t) \\ P(10, t) \\ P(11, t) \end{pmatrix} \quad (5.76)$$

donde $\bar{f}(t) = P(01, t) + P(10, t) + 2P(11, t)$. Notemos que la primera columna es cero inmediatamente, esto reduce la dimensión del espacio. Esto es razonable pues esperamos que la cadena 00 no contribuya a las proporciones dado su valor de aptitud nulo.

Los valores propios junto con los vectores propios se muestran a continuación: $\lambda_0 = 0$ tiene multiplicidad 1 y asociado el vector $[1, 0, 0, 0]$, $\lambda_1 = -3p_m + \frac{3}{2} +$

$2p_m^2 - \frac{1}{2}\beta_0$ que tiene asociado al vector $[-\frac{(p_m^2 - p_m + \frac{1}{2} + \frac{1}{2}\beta_0)}{(1-2p_m + p_m^2)}, \frac{1}{2}\frac{(-p_m + \frac{1}{2} + \frac{1}{2}\beta_0)}{p_m(-1+p_m)}, \frac{1}{2}\frac{(-p_m + \frac{1}{2} + \frac{1}{2}\beta_0)}{p_m(-1+p_m)}, 1]$, $\lambda_2 = -3p_m + \frac{3}{2} + 2p_m^2 + \frac{1}{2}\beta_0$ de multiplicidad 1 y el vector $[-\frac{(p_m^2 - p_m + \frac{1}{2} - \frac{1}{2}\beta_0)}{(1-2p_m + p_m^2)}, \frac{1}{2}\frac{(-p_m + \frac{1}{2} - \frac{1}{2}\beta_0)}{p_m(-1+p_m)}, \frac{1}{2}\frac{(-p_m + \frac{1}{2} - \frac{1}{2}\beta_0)}{p_m(-1+p_m)}, 1]$ y $\lambda_3 = 1 - 2p_m$ de multiplicidad 1, y tiene asociado el vector $[0, 1, -1, 0]$ con $\beta_0 = \sqrt{16p_m^4 - 32p_m^3 + 20p_m^2 - 4p_m + 1}$. Por lo tanto la evolución de las proporciones está dada por:

$$P(t+1) = \frac{1}{\bar{f}(t)} \bar{Q} \bar{D} \bar{Q} P(t) \quad (5.77)$$

Con:

$$P(t) = \begin{pmatrix} P(00, t) \\ P(01, t) \\ P(10, t) \\ P(11, t) \end{pmatrix} \quad (5.78)$$

el vector de proporciones,

$$\bar{Q} = \begin{pmatrix} 1 & -\frac{p_m^2 - p_m + 1/2 + \beta_0/2}{1 - 2p_m + p_m^2} & -\frac{p_m^2 - p_m + 1/2 - \beta_0/2}{1 - 2p_m + p_m^2} & 0 \\ 0 & 1/2 \frac{-p_m + 1/2 + \beta_0/2}{p_m(-1+p_m)} & 1/2 \frac{-p_m + 1/2 - \beta_0/2}{p_m(-1+p_m)} & -1 \\ 0 & 1/2 \frac{-p_m + 1/2 + \beta_0/2}{p_m(-1+p_m)} & 1/2 \frac{-p_m + 1/2 - \beta_0/2}{p_m(-1+p_m)} & 1 \\ 0 & 1 & 1 & 0 \end{pmatrix} \quad (5.79)$$

la matriz de transición de base,

$$\bar{D} = \begin{pmatrix} \lambda_0 & 0 & 0 & 0 \\ 0 & \lambda_1 & 0 & 0 \\ 0 & 0 & \lambda_2 & 0 \\ 0 & 0 & 0 & \lambda_3 \end{pmatrix} \quad (5.80)$$

y $\bar{f}(t) = 2P(11, t) + P(10, t) + P(01, t)$

Si en la ecuación (5.77) multiplicamos ambos lados por \bar{Q} a la izquierda tenemos:

$$\bar{Q} P(t+1) = \frac{1}{\bar{f}(t)} \bar{D} \bar{Q} P(t) \quad (5.81)$$

Ahora hagamos el siguiente cambio de base:

$$\begin{pmatrix} \alpha_0(t) \\ \alpha_1(t) \\ \alpha_2(t) \\ \alpha_3(t) \end{pmatrix} = \alpha(t) = \bar{Q} P \quad (5.82)$$

Tenemos entonces que:

$$\alpha(t+1) = \frac{1}{\bar{f}(t)} \bar{D} \alpha(t) \quad (5.83)$$

y dado que D es diagonal fácilmente podemos iterar el sistema obteniendo:

$$\alpha(t) = \left(\prod_{i=0}^t \frac{1}{f(i)} \right) \overline{D}^t \alpha(0) \quad (5.84)$$

Si tomamos cocientes podemos eliminar la aptitud promedio y obtenemos el siguiente sistema:

$$\alpha_0(t) = 0 \quad (5.85)$$

$$\frac{\alpha_1(t)}{\alpha_2(t)} = \left(\frac{\lambda_1}{\lambda_2} \right)^t \frac{\alpha_1(0)}{\alpha_2(0)} \quad (5.86)$$

$$\frac{\alpha_3(t)}{\alpha_2(t)} = \left(\frac{\lambda_3}{\lambda_2} \right)^t \frac{\alpha_3(0)}{\alpha_2(0)} \quad (5.87)$$

$$1 = \alpha_0 + \alpha_1(t)(\gamma_1 + 2\gamma_3 + 1) + \alpha_2(t)(\gamma_2 + 2\gamma_4 + 1) \quad (5.88)$$

La última ecuación se obtiene de expresar la relación de conservación de probabilidad ($P(00, t) + P(01, t) + P(10, t) + P(11, t) = 1$) en la nueva base. Con $\gamma_1 = -\frac{p_m^2 - p_m + 1/2 + \beta_0/2}{1 - 2p_m + p_m^2}$, $\gamma_2 = -\frac{p_m^2 - p_m + 1/2 - \beta_0/2}{1 - 2p_m + p_m^2}$, $\gamma_3 = \frac{-p_m + 1/2 + \beta_0/2}{2p_m(-1 + p_m)}$ y $\gamma_4 = \frac{-p_m + 1/2 - \beta_0/2}{2p_m(-1 + p_m)}$

Si iteramos este sistema y resolviendo para $\alpha_0(t)$, $\alpha_1(t)$, $\alpha_2(t)$ y $\alpha_3(t)$ tenemos:

$$\alpha_0(t) = 0 \quad (5.89)$$

$$\alpha_1(t) = \left(\frac{\lambda_1}{\lambda_2} \right)^t \left(\frac{\alpha_1(0)}{\alpha_2(0)} \right) \alpha_2(t) \quad (5.90)$$

$$\alpha_2(t) = \left(\left(\frac{\lambda_1}{\lambda_2} \right)^t \left(\frac{\alpha_1(0)}{\alpha_2(0)} \right) (\gamma_1 + 2\gamma_3 + 1) + \gamma_2 + 2\gamma_4 + 1 \right)^{-1} \quad (5.91)$$

$$\alpha_3(t) = \left(\frac{\lambda_3}{\lambda_2} \right)^t \left(\frac{\alpha_3(0)}{\alpha_2(0)} \right) \alpha_2(t) \quad (5.92)$$

Una vez resuelto para la base $\{\alpha_1(t), \alpha_2(t), \alpha_3(t) \text{ y } \alpha_4(t)\}$, solo nos resta hacer la transformación inversa para obtener la solución al sistema de ecuaciones en términos de la base original ($\mathbf{P}(t) = \overline{\mathbf{Q}}\alpha(t)$).

$$P(00, t) = \left(\gamma_2 + \gamma_1 \left(\frac{\lambda_1}{\lambda_2} \right)^t \left(\frac{\alpha_1(0)}{\alpha_2(0)} \right) \right) \alpha_2(t) \quad (5.93)$$

$$P(01, t) = \left(\gamma_4 + \gamma_3 \left(\frac{\lambda_1}{\lambda_2} \right)^t \left(\frac{\alpha_1(0)}{\alpha_2(0)} \right) - \left(\frac{\lambda_3}{\lambda_2} \right)^t \left(\frac{\alpha_3(0)}{\alpha_2(0)} \right) \right) \alpha_2(t) \quad (5.94)$$

$$P(10, t) = \left(\gamma_4 + \gamma_3 \left(\frac{\lambda_1}{\lambda_2} \right)^t \left(\frac{\alpha_1(0)}{\alpha_2(0)} \right) + \left(\frac{\lambda_3}{\lambda_2} \right)^t \left(\frac{\alpha_3(0)}{\alpha_2(0)} \right) \right) \alpha_2(t) \quad (5.95)$$

$$P(11, t) = \left(1 + \left(\frac{\lambda_1}{\lambda_2} \right)^t \left(\frac{\alpha_1(0)}{\alpha_2(0)} \right) \right) \alpha_2(t) \quad (5.96)$$

$$(5.97)$$

Para examinar el comportamiento de los puntos fijos, tomamos $\lim_{t \rightarrow \infty}$. Los cocientes $\frac{\lambda_2}{\lambda_3}$ y $\frac{\lambda_4}{\lambda_3}$ se comportan como en la figura 5.8, es decir, su valor absoluto es menor que 1 para $p_m \in [0, 1)$. Por lo tanto, al tomar el límite para t grande se anulan y tenemos:

$$\lim_{t \rightarrow \infty} P(00, t) = \frac{p_m(4p_m^2 - 4p_m + \beta_0)}{2} \quad (5.98)$$

$$\lim_{t \rightarrow \infty} P(01, t) = \frac{\beta_0}{2} \left(p_m - \frac{1}{2}\right) - 3p_m^2 + p_m + 2p_m^3 + \frac{1}{4} \quad (5.99)$$

$$\lim_{t \rightarrow \infty} P(10, t) = \frac{\beta_0}{2} \left(p_m - \frac{1}{2}\right) - 3p_m^2 + p_m + 2p_m^3 + \frac{1}{4} \quad (5.100)$$

$$\lim_{t \rightarrow \infty} P(11, t) = \frac{(1-p_m)(4p_m - 4p_m^2 - 1 - \beta_0)}{2} \quad (5.101)$$

$$(5.102)$$

Los resultados pueden observarse en la figura 5.9. Claramente notamos un punto de inflexión en $\frac{1}{2}$ y, como era esperado, las soluciones para las cuerdas 01 y 10 son idénticas. Estas mismas cuerdas predominan conforme aumenta p_m pues 11 se convierte en 00 y queda eliminada eventualmente mientras que 01 y 10 se convierten una en la otra y se preservan. En [5] se muestran los resultados de aplicar un cambio de escala a esas ecuaciones de evolución.

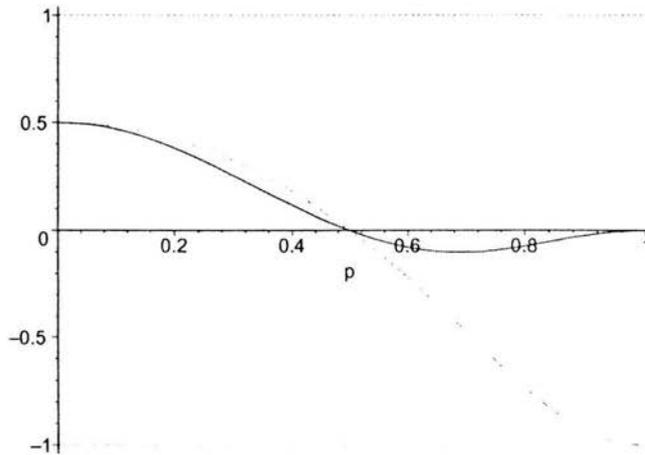


Figura 5.8: Selección y mutación 2 bits: Cociente de los valores característicos $\frac{\lambda_2}{\lambda_3}$ (línea continua) y $\frac{\lambda_4}{\lambda_3}$ (línea punteada) como función de p_m . Nótese que para $p_m \in [0, 1)$ el valor absoluto de estos cocientes es menor que 1 (líneas discontinuas)

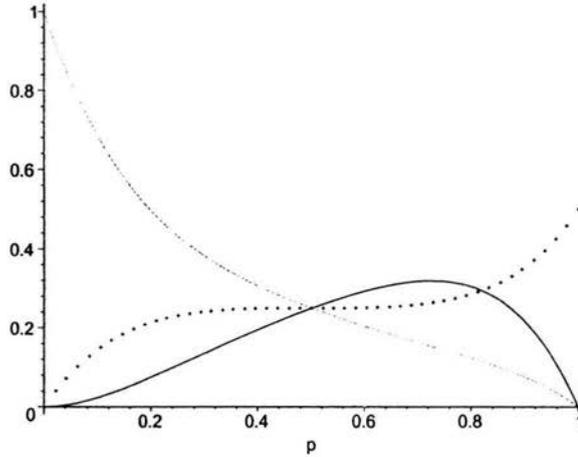


Figura 5.9: Selección y mutación para 2 bits contando 1s: Comportamiento asintótico como función de p_m de $P(00, t)$ (línea continua), $P(01, t) = P(10, t)$ (cruces) y $P(11, t)$ (línea discontinua)

5.5.3. Modelo de Eigen

Si suponemos que 11 es la secuencia maestra tenemos:

$$\begin{pmatrix} P(00, t+1) \\ P(01, t+1) \\ P(10, t+1) \\ P(11, t+1) \end{pmatrix} = \begin{pmatrix} (1-p_m)^2 & (1-p_m)p_m & (1-p_m)p_m & p_m^2 \\ (1-p_m)p_m & (1-p_m)^2 & p_m^2 & (1-p_m)p_m \\ (1-p_m)p_m & p_m^2 & (1-p_m)^2 & (1-p_m)p_m \\ p_m^2 & (1-p_m)p_m & (1-p_m)p_m & (1-p_m)^2 \end{pmatrix} \times \frac{1}{\bar{f}(t)} \begin{pmatrix} f_0 & 0 & 0 & 0 \\ 0 & f_0 & 0 & 0 \\ 0 & 0 & f_0 & 0 \\ 0 & 0 & 0 & f_{sm} \end{pmatrix} \begin{pmatrix} P(00, t) \\ P(01, t) \\ P(10, t) \\ P(11, t) \end{pmatrix} \quad (5.103)$$

Con $\bar{f}(t) = f_0(P(00, t) + P(01, t) + P(10, t)) + f_{sm}P(11, t)$. Primeramente obtenemos los valores propios $(\lambda_i, i \in \{0, 1, 2, 3\})$ y vectores propios con los que formamos las matrices \bar{D} y \bar{Q} respectivamente, donde

$$\bar{D} = \begin{pmatrix} \lambda_0 & 0 & 0 & 0 \\ 0 & \lambda_1 & 0 & 0 \\ 0 & 0 & \lambda_2 & 0 \\ 0 & 0 & 0 & \lambda_3 \end{pmatrix} \quad (5.104)$$

y los vectores propios son las columnas de \bar{Q} , la matriz de transición de base. Tenemos la ecuación:

$$\mathbf{P}(t+1) = \frac{1}{f(t)} \overline{\mathbf{Q}} \mathbf{D} \overline{\mathbf{Q}} \mathbf{P}(t) \quad (5.105)$$

Con:

$$\mathbf{P}(t) = \begin{pmatrix} P(00,t) \\ P(01,t) \\ P(10,t) \\ P(11,t) \end{pmatrix} \quad (5.106)$$

el vector de proporciones,

Si en la ecuación (5.105) multiplicamos ambos lados por $\overline{\mathbf{Q}}$ a la izquierda tenemos:

$$\overline{\mathbf{Q}} \mathbf{P}(t+1) = \frac{1}{f(t)} \overline{\mathbf{D}} \overline{\mathbf{Q}} \mathbf{P}(t) \quad (5.107)$$

Ahora hagamos el siguiente cambio de base:

$$\begin{pmatrix} \alpha_0(t) \\ \alpha_1(t) \\ \alpha_2(t) \\ \alpha_3(t) \end{pmatrix} = \boldsymbol{\alpha}(t) = \overline{\mathbf{Q}} \mathbf{P} \quad (5.108)$$

Tenemos entonces que:

$$\boldsymbol{\alpha}(t+1) = \frac{1}{f(t)} \overline{\mathbf{D}} \boldsymbol{\alpha}(t) \quad (5.109)$$

y dado que D es diagonal fácilmente podemos iterar el sistema obteniendo:

$$\boldsymbol{\alpha}(t) = \left(\prod_{i=0}^{t-1} \frac{1}{f(i)} \right) \overline{\mathbf{D}}^t \boldsymbol{\alpha}(0) \quad (5.110)$$

Los valores característicos ($\lambda_i, i \in \{1, 2, 3, 4\}$) son polinomios en la variable p_m , elegimos uno de ellos digamos λ_k tal que no se anula en el intervalo $[0, 1]$. Entonces, si tomamos los cocientes $\frac{\lambda_i}{\lambda_k}, i \in \{1, 2, 3, 4\}, i \neq k$, podemos eliminar la aptitud promedio del sistema y nos quedamos con 3 ecuaciones y 4 variables. La cuarta ecuación la obtenemos aplicando el cambio de base a la ecuación de conservación de probabilidad. Ahora, si iteramos este nuevo sistema y resolvemos para $\alpha_0(t), \alpha_1(t), \alpha_2(t)$ y $\alpha_3(t)$, solo nos resta hacer la transformación inversa para obtener la solución al sistema de ecuaciones en términos de la base original ($\mathbf{P}(t) = \overline{\mathbf{Q}} \boldsymbol{\alpha}(t)$). En [5] se muestran los resultados de aplicar un cambio de escala a esas ecuaciones de evolución.

5.6. Soluciones para Selección y cruza 2 bits

En el caso de cruza, comenzamos con el caso de 2 bits pues es el primero en el que tiene sentido aplicar este operador. De la ecuación (3.14) tenemos:

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

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

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

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

5.6.1. Escenario plano

Las ecuaciones en términos de cuerdas no son muy útiles, sin embargo, si expresamos las ecuaciones de cruza en términos de esquemas:

$$P(00, t + 1) = (1 - p_c)P(00, t) + p_cP(0*, t)P(*0, t) \quad (5.111)$$

$$P(01, t + 1) = (1 - p_c)P(01, t) + p_cP(0*, t)P(*1, t) \quad (5.112)$$

$$P(10, t + 1) = (1 - p_c)P(10, t) + p_cP(1*, t)P(*0, t) \quad (5.113)$$

$$P(11, t + 1) = (1 - p_c)P(11, t) + p_cP(1*, t)P(*1, t) \quad (5.114)$$

Dada la simetría consideremos solo la cadena ij , $i, j \in \{0, 1\}$, si iteramos obtenemos la siguiente solución:

$$\begin{aligned} P(ij, t) &= (1 - p_c)^t P(ij, t) + p_c \left(\sum_{k=0}^{t-1} (1 - p_c)^k \right) P(i*, 0)P(*j, 0) \\ &= (1 - p_c)^t P(ij, t) + (1 - (1 - p_c)^{t-1}) P(i*, 0)P(*j, 0) \end{aligned} \quad (5.115)$$

Debemos recordar que los 1-esquemas no son afectados por el operador de cruza, es decir que si tenemos ξ_1 tal que $N_2(\xi_1) = 1$ entonces $P(\xi_1, t) = P(\xi_1, 0)$. De lo anterior podemos obtener fácilmente el comportamiento en el límite para t grande

$$\lim_{t \rightarrow \infty} P(ij, t) = P(i*, 0)P(*j, 0) \lim_{t \rightarrow \infty} (1 - (1 - p_c)^{t-1}) = P(i*, 0)P(*j, 0) \quad (5.116)$$

Notemos que en términos de esquemas el punto fijo es único y es básicamente el teorema de Geiringer [8] para dos bits en el contexto de distribuciones de esquemas y cruza de un punto. Nuevamente si asumimos una distribución aleatoria, el límite de las proporciones de cada una de las cadenas será igual.

Notemos también que el término de destrucción ($(1 - p_c)^t P(ij, t)$) que es el que se considera en el teorema de esquemas, es poco significativo asintóticamente en comparación con el término de construcción. De aquí la cota del teorema de esquemas es pobre.

En la sección A.4 del apéndice de aproximaciones tomando límite de tiempo continuo se muestran argumentos de convergencia para estas ecuaciones.

5.6.2. Escenario de contando 1's

Esta vez contando 1's y usando selección proporcional como antes, tenemos:

$$\begin{aligned}
 P(00, t+1) &= \frac{f(00)}{\bar{f}(t)} P(00, t) - \frac{p_c}{\bar{f}^2(t)} [f(00)P(00, t)f(11)P(11, t) - f(01)P(01, t)f(10)P(10, t)] \\
 P(01, t+1) &= \frac{f(01)}{\bar{f}(t)} P(01, t) - \frac{p_c}{\bar{f}^2(t)} [f(01)P(01, t)f(01)P(10, t) - f(00)P(00, t)f(11)P(11, t)] \\
 P(10, t+1) &= \frac{f(10)}{\bar{f}(t)} P(10, t) - \frac{p_c}{\bar{f}^2(t)} [f(10)P(10, t)f(01)P(01, t) - f(00)P(00, t)f(11)P(11, t)] \\
 P(11, t+1) &= \frac{f(11)}{\bar{f}(t)} P(11, t) - \frac{p_c}{\bar{f}^2(t)} [f(11)P(11, t)f(00)P(00, t) - f(01)P(01, t)f(10)P(10, t)]
 \end{aligned}$$

Considerando el valor real de la aptitud, como antes ($\alpha + w^H(C_i)$) tenemos:

$$\begin{aligned}
 P(00, t+1) &= \frac{\alpha}{\bar{f}(t)} P(00, t) - \frac{p_c}{\bar{f}^2(t)} [\alpha(2 + \alpha)P(00, t)P(11, t) - (1 + \alpha)^2 P(01, t)P(10, t)] \\
 P(01, t+1) &= \frac{(1 + \alpha)}{\bar{f}(t)} P(01, t) - \frac{p_c}{\bar{f}^2(t)} [(1 + \alpha)^2 P(01, t)P(10, t) - \alpha(2 + \alpha)P(00, t)P(11, t)] \\
 P(10, t+1) &= \frac{(1 + \alpha)}{\bar{f}(t)} P(10, t) - \frac{p_c}{\bar{f}^2(t)} [(1 + \alpha)^2 P(10, t)P(01, t) - \alpha(2 + \alpha)P(00, t)P(11, t)] \\
 P(11, t+1) &= \frac{(2 + \alpha)}{\bar{f}(t)} P(11, t) - \frac{p_c}{\bar{f}^2(t)} [\alpha(2 + \alpha)P(11, t)P(00, t) - (1 + \alpha)^2 P(01, t)P(10, t)]
 \end{aligned}$$

En el caso especial cuando $\alpha \rightarrow 0$

$$\begin{aligned}
 P(00, t+1) &= \frac{p_c}{\bar{f}^2(t)} [P(01, t)P(10, t)] \\
 P(01, t+1) &= \frac{1}{\bar{f}(t)} P(01, t) - \frac{p_c}{\bar{f}^2(t)} [P(01, t)P(10, t)] \\
 P(10, t+1) &= \frac{1}{\bar{f}(t)} P(10, t) - \frac{p_c}{\bar{f}^2(t)} [P(10, t)P(01, t)] \\
 P(11, t+1) &= \frac{2}{\bar{f}(t)} P(11, t) + \frac{p_c}{\bar{f}^2(t)} [P(01, t)P(10, t)]
 \end{aligned}$$

Notemos que la fuente principal de cadenas es la cruza de 01 con 10 pues no podemos contar con la cadena 00 para producir otras. Si restamos la segunda y tercera ecuaciones observamos que $P(01, t) = P(10, t)$ para t grande pues $\bar{f}(t) > P(01, t) - P(10, t)$. Por el mismo argumento, si sustituimos este resultado en las ecuaciones 2 y 3 por separado, también notamos que $P(01, t) = P(10, t) = 0$ para $t \rightarrow \infty$. Y de esto $\lim_{t \rightarrow \infty} P(11, t) = 1$. En [5] se habla más a detalle y muestra una cota para $\Delta(t)$.

5.6.3. Modelo de Eigen

Si suponemos 11 como la secuencia maestra, las ecuaciones son las siguientes:

$$\begin{aligned}
 P(00, t+1) &= \frac{f_0}{\bar{f}(t)} P(00, t) - \frac{p_c}{\bar{f}^2(t)} [f_0 f_{sm} P(00, t) P(11, t) - f_0^2 P(01, t) P(10, t)] \\
 P(01, t+1) &= \frac{f_0}{\bar{f}(t)} P(01, t) + \frac{p_c}{\bar{f}^2(t)} [f_0 f_{sm} P(00, t) P(11, t) - f_0^2 P(01, t) P(10, t)] \\
 P(10, t+1) &= \frac{f_0}{\bar{f}(t)} P(10, t) + \frac{p_c}{\bar{f}^2(t)} [f_0 f_{sm} P(00, t) P(11, t) - f_0^2 P(01, t) P(10, t)] \\
 P(11, t+1) &= \frac{f_{sm}}{\bar{f}(t)} P(11, t) - \frac{p_c}{\bar{f}^2(t)} [f_0 f_{sm} P(00, t) P(11, t) - f_0^2 P(01, t) P(10, t)]
 \end{aligned}$$

Por los mismos argumentos del caso de contando unos, las proporciones de las cuerdas 01 y 10 se igualan a cero para t grande. De igual manera $P(00, t)$ converge a cero, por lo que la secuencia maestra predomina.

5.7. Soluciones para Selección y cruza 3 bits

5.7.1. Escenario plano

Llamemos ijk a una cuerda de 3 bits cualquiera con $i, j, k \in \{0, 1\}$, si iteramos usando al ecuación en términos de esquemas para cruza tenemos:

$$P(ijk, t+1) = (1 - p_c) P(ijk, t) + \frac{p_c}{2} [P(i**0) P(*jk, t) + P(ij*, t) P(**k, 0)] \quad (5.117)$$

Donde

$$P(*jk, t) = \left(1 - \frac{p_c}{2}\right) P(*jk, t-1) + \frac{p_c}{2} P(*j*, 0) P(**k, 0) \quad (5.118)$$

$$P(ij*, t) = \left(1 - \frac{p_c}{2}\right) P(ij*, t-1) + \frac{p_c}{2} P(i**0) P(*j*, 0) \quad (5.119)$$

Notamos que estas ecuaciones tienen la misma forma que las de 2 bits y por lo tanto unas soluciones tienen la siguiente forma:

$$P(ij*, t) = \left(1 - \frac{p_c}{2}\right)^t P(ij*, t) + \left(1 - \left(1 - \frac{p_c}{2}\right)^{t-1}\right) P(i**0) P(*j*, 0) \quad (5.120)$$

$$P(*jk, t) = \left(1 - \frac{p_c}{2}\right)^t P(*jk, t) + \left(1 - \left(1 - \frac{p_c}{2}\right)^{t-1}\right) P(*j*, 0) P(**k, 0) \quad (5.121)$$

Si sustituimos estas soluciones tenemos

$$\begin{aligned}
 P(ijk, t+1) &= (1 - p_c) P(ijk, t) + \\
 &\frac{p_c}{2} \left[P(i**0) \left(\left(1 - \frac{p_c}{2}\right)^t P(*jk, t) + \left(1 - \left(1 - \frac{p_c}{2}\right)^{t-1}\right) P(*j*, 0) P(**k, 0) \right) + \right. \\
 &\quad \left. \left(\left(1 - \frac{p_c}{2}\right)^t P(ij*, t) + \left(1 - \left(1 - \frac{p_c}{2}\right)^{t-1}\right) P(i**0) P(*j*, 0) \right) P(**k, 0) \right] \quad (5.122)
 \end{aligned}$$

Cuya solución a la vez es:

$$\begin{aligned}
 P(ijk, t+1) &= (1-p_c)^t P(ijk, t) + \left(\frac{1 - (1 - \frac{p_c}{2})^{t-1}}{2} \right) \\
 &\left[P(i^{**}, 0) \left((1-p_c)^t P(*jk, t) + \left(1 - (1 - \frac{p_c}{2})^{t-1} \right) P(*j^*, 0) P(**k, 0) \right) + \right. \\
 &\left. \left((1-p_c)^t P(ij^*, t) + \left(1 - (1 - \frac{p_c}{2})^{t-1} \right) P(i^{**}, 0) P(*j^*, 0) \right) P(**k, 0) \right]
 \end{aligned}
 \tag{5.123}$$

Tomando límite para t grande obtenemos

$$\lim_{t \rightarrow \infty} P(ijk, t) = P(i^{**}, t) P(*j^*, t) P(**k, t) \tag{5.124}$$

Nuevamente encontramos que en el contexto de esquemas hay un único punto fijo y coincide con el teorema de Geiringer [8] para tres bits. Podemos obtener el mismo resultado tomando el límite de tiempo continuo simplemente como se muestra en el apéndice A.5

**ESTA TESIS NO SALE
DE LA BIBLIOTECA**

Capítulo 6

Soluciones analíticas, soluciones numéricas

En el capítulo anterior consideramos algunos modelos de juguete y adquirimos intuición sobre la dinámica de los AGs tanto a nivel de cuerdas como a nivel de esquemas y otros cambio de escala. Ahora podemos extender esta intuición a escenarios más generales y examinar el comportamiento de las ecuaciones que gobiernan el comportamiento de los AGs.

6.1. Ecuación general formal

Podemos expresar los resultados obtenidos en la sección 3.3 en notación matricial como sigue:

$$\mathbf{P}(t+1) = \overline{\mathbf{W}}\mathbf{P}_c(t) \quad (6.1)$$

donde la matriz de mutación $\overline{\mathbf{W}}$ es real simétrica y tiene entradas $\overline{W}_{ij} = p_m^{d''(i,j)}(1-p_m)^{N-d''(i,j)}$. Restringiendo nuestra atención al caso de esquemas de selección lineales en $P(C_i, t)$, $\mathbf{P}_c(t)$ se puede escribir como

$$\mathbf{P}_c(t) = \overline{\mathbf{F}}(t)\mathbf{P}(t) + \sum_{m=1}^{2^N} p_c(m)\mathbf{j}(m, t) \quad (6.2)$$

donde la matriz de selección y destrucción por cruza, $\overline{\mathbf{F}}(t)$, es diagonal ($\overline{F}_{ii}(t) = \frac{f(C_i)}{f(t)}(1-p_c)$). La matriz $\mathbf{j}(m, t) = P'(C_i^s(m), t)P'(C_i^c(m), t)$ y definiendo la matriz de selección, destrucción por cruza y mutación $\overline{\mathbf{W}}_s(t) = \overline{\mathbf{W}}\overline{\mathbf{F}}(t)$ tenemos

$$\mathbf{P}(t+1) = \overline{\mathbf{W}}_s(t)\mathbf{P}(t) + \sum_{m=1}^{2^N} p_c(m)\overline{\mathbf{W}}\mathbf{j}(m, t) \quad (6.3)$$

La interpretación de esta ecuación es que $\mathbf{j}(m, t)$ es una “fuente” que crea cadenas (o esquemas) juntando bloques constructores, mientras el primer término del lado derecho nos dice como las cadenas son propagadas a la siguiente generación. El efecto destructivo de la cruza renormaliza la aptitud de las cadenas. La solución iterada a la ecuación (6.3) es, como se muestra en [23]:

$$\mathbf{P}(t) = \prod_{n=0}^{t-1} (\overline{\mathbf{W}}_s(n)) \mathbf{P}(0) + \sum_{m=1}^{2^N} p_c(m) \sum_{n=0}^{t-1} \prod_{i=n}^{t-1} \overline{\mathbf{W}}_s(i) \overline{\mathbf{W}} \mathbf{j}(m, n) \quad (6.4)$$

Siendo $\overline{\mathbf{W}}$ la matriz de mutación; es real, simétrica y diagonalizable vía una transformación de similaridad $\overline{\mathbf{Q}}$ tal que $\overline{\mathbf{Q}} \overline{\mathbf{W}} \overline{\mathbf{Q}}^{-1} = \overline{\mathbf{D}}$ donde $\overline{\mathbf{D}}$ es diagonal. La interpretación de (6.3) se sigue naturalmente de la de (6.4). Considerando primero el caso sin mutación, el primer término del lado derecho nos da la probabilidad de propagar una cadena o esquema del tiempo cero a t sin ser destruida por la cruza. En otras palabras $\prod_{n=0}^{t-1} \overline{\mathbf{D}}_s(n)$ es la función de Green o propagador de \mathbf{P} . En el segundo término, $\mathbf{j}(m, n)$ cada elemento está asociado con la creación de una cadena o esquema al tiempo n vía la yuxtaposición de dos bloques constructores asociados con la máscara m . El factor $\prod_{i=n}^{t-1} \overline{\mathbf{D}}_s(i)$ es la probabilidad de propagar la cadena o esquema resultante sin destrucción por cruza desde su creación al tiempo n al tiempo t . La suma sobre máscaras y n es simplemente la suma sobre todos los eventos de creación en la dinámica.

Observemos que esta solución tiene una forma jerárquica muy clara, donde la raíz es una cadena al tiempo t , sus descendientes directos son las probabilidades de juntar de sus bloques constructores al tiempo n , cuyos descendientes a su vez son sus propios bloques constructores, etc. Para expresar esta estructura hacemos uso de una representación diagramática. En la figura 6.1 se muestran las 2 posibilidades para el caso de 2 bits. Para el caso de cuatro bits se muestran en la fig. 6.2. En estos diagramas podemos observar la forma en la que se unen los bloques constructores para dar origen a la cadena $ijkl$: la forma más directa es que exista desde el inicio y se conserve en esa forma. Sin embargo lo que sucede con mayor frecuencia es que se unen los bloques con el tiempo propagándose hasta encontrar otro bloque que los complementa para formar la solución.

6.2. Solución para la ecuación de selección

Si ahora consideramos solo selección proporcional tenemos

$$P(C_i, t+1) = \frac{f(C_i)}{\bar{f}(t)} P(C_i, t) \quad (6.5)$$

por definición

$$\bar{f}(t) = \sum_{C_i} f(C_i) \frac{n(C_i, t)}{n} \quad (6.6)$$

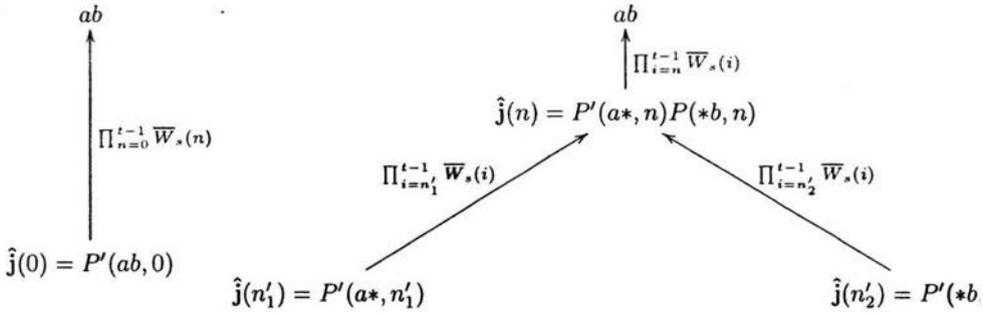


Figura 6.1: Solución diagramática para 2 bits: A la izquierda vemos la cadena ab propagada desde el tiempo $t = 0$ con fuerza $\prod_{n=0}^{t-1} \bar{W}_s(n)$. En el lado derecho vemos su construcción al tiempo n por medio de los bloques constructores a^* y $*b$ y después su propagación con fuerza $\prod_{i=n}^{t-1} \bar{W}_s(i)$

Así es que usamos el siguiente cambio de variable:

$$P(C_i, t) = \frac{n(C_i, t)}{n} = \frac{y(C_i, t)}{\sum_{C_i} y(C_i, t)} \quad (6.7)$$

y tenemos:

$$\frac{y(C_i, t+1)}{\sum_{C_i} y(C_i, t+1)} = \frac{f(C_i)y(C_i, t)}{\sum_{C_i} y(C_i, t)} \frac{1}{\sum_{C_i} \frac{f(C_i)y(C_i, t)}{\sum_{C_i} y(C_i, t)}} \quad (6.8)$$

Y podemos cancelar $\sum_{C_i} y(C_i, t)$ obteniendo así:

$$\frac{y(C_i, t+1)}{\sum_{C_i} y(C_i, t+1)} = \frac{f(C_i)y(C_i, t)}{\sum_{C_i} f(C_i)y(C_i, t)} \quad (6.9)$$

que tiene a $y(C_i, t+1) = y(C_i, t)f(C_i)$ por solución. Entonces si iteramos sobre el tiempo obtenemos la solución exacta para selección:

$$P(C_i, t) = \frac{f^t(C_i)P(C_i, 0)}{\sum_{C_j} f^t(C_j)P(C_j, 0)} \quad (6.10)$$

A continuación mostramos la solución iterada para selección, pero esta vez haciendo el cambio de escala de esquemas.

Teorema 6.2.1 Si $P(\xi_i, t)$ es la proporción del esquema ξ al tiempo t entonces, para un esquema de selección proporcional tenemos

$$P(\xi, t) = \frac{f^t(\xi, 0)P(\xi, 0)}{\sum_j f^t(C_j)P(C_j, 0)} \quad (6.11)$$

Demostración

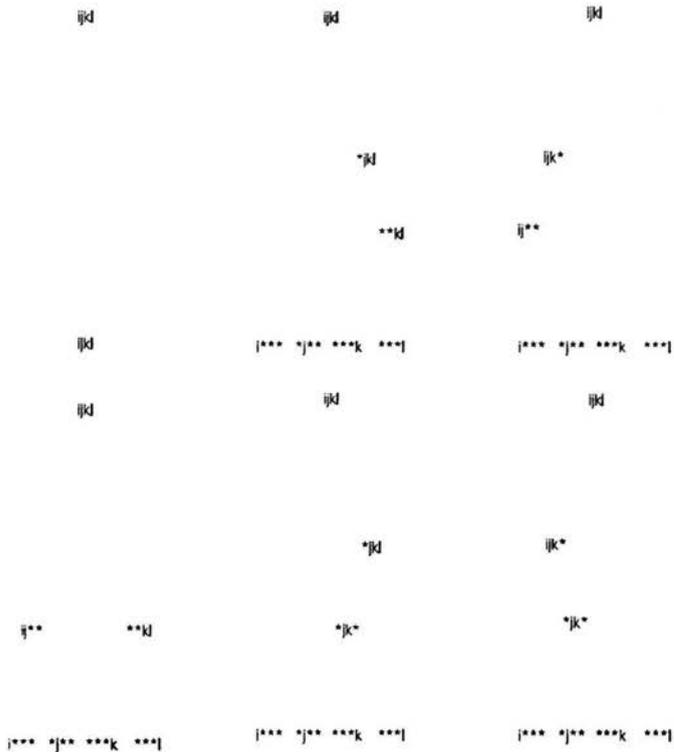


Figura 6.2: Diagramas de Feynman para 4 bits

Por inducción. Para $t = 0$ es claro pues

$$P(\xi, 0) = \frac{P(\xi, 0)}{\sum_j f^0(C_j)P(C_j, 0)} = P(\xi, 0) \quad (6.12)$$

Ahora suponemos que el resultado es cierto para $t - 1$, es decir

$$P(\xi, t - 1) = \frac{\left(\prod_{l=0}^{t-1} f(\xi, l)\right) P(\xi, 0)}{\sum_j f^{t-1}(C_j)P(C_j, 0)} \quad (6.13)$$

Para el tiempo t tenemos por definición que

$$\sum_{C_k \in \xi} P(C_k, t) = P(\xi, t) = \frac{f(\xi, t)P(\xi, t-1)}{\sum_j f(C_j)P(C_j, t-1)} \quad (6.14)$$

En el numerador sustituimos la hipótesis de inducción:

$$f(\xi, t) \left(\frac{\left(\prod_{l=0}^{t-1} f(\xi, l) \right) P(\xi, 0)}{\sum_{j'} f^{t-1}(C_{j'}) P(C_{j'}, 0)} \right) \quad (6.15)$$

y en el denominador hacemos uso del resultado anterior

$$\sum_j f(C_j) \left(\frac{f^{t-1}(C_j) P(C_j)}{\sum_{j'} f^{t-1}(C_{j'}) P(C_{j'}, 0)} \right) \quad (6.16)$$

Podemos cancelar $\sum_{j'} f^{t-1}(C_{j'}) P(C_{j'}, 0)$ de ambos y tenemos el resultado.

Notemos que la solución para la ecuación de selección haciendo el cambio de escala de esquemas tiene la misma forma que antes, aunque hay que recalcar la dependencia en el tiempo de la función de aptitud para esquemas.

6.3. Solución para la ecuación de mutación

En [20] se da una demostración de cómo la dinámica de mutación puede ser descrita en términos de la base de vectores propios. Este resultado es la generalización lógica de nuestras soluciones para el caso de pocos bits.

Sea $\bar{\mathbf{A}}$ la matriz de transición para el operador de mutación para n bits. Como antes, encontramos sus valores propios $(\lambda_1 \dots \lambda_{2^n})$ y vectores propios. Para formar la matriz $\bar{\mathbf{Q}}$ tomamos los vectores propios como columnas, dado que tenemos 2^n vectores propios linealmente independientes para un espacio de dimensión 2^N , la matriz es cuadrada, invertible y diagonaliza a $\bar{\mathbf{A}}$ de la siguiente manera: $\tilde{\mathbf{Q}}\bar{\mathbf{A}}\bar{\mathbf{Q}} = \text{diag}\{\lambda_1, \dots, \lambda_{2^n}\}$, con $\tilde{\mathbf{Q}}$ la matriz inversa de $\bar{\mathbf{Q}}$. Entonces transformando la ecuación de evolución de cuerdas en notación matricial tenemos:

$$\begin{aligned} \tilde{\mathbf{Q}}\mathbf{P}(t+1) &= \frac{1}{\mathbf{f}^T\mathbf{P}(t)} \tilde{\mathbf{Q}}\bar{\mathbf{A}}\text{diag}(\mathbf{f})\mathbf{P}(t) \\ &= \frac{1}{\mathbf{f}^T\mathbf{P}(t)} \tilde{\mathbf{Q}}\bar{\mathbf{Q}}\text{diag}(\lambda)\tilde{\mathbf{Q}}\mathbf{P}(t) \\ &= \frac{1}{\mathbf{f}^T\mathbf{P}(t)} \text{diag}(\lambda)\tilde{\mathbf{Q}}\mathbf{P}(t) \end{aligned}$$

Para continuar, usamos una expresión alternativa para $\mathbf{f}^T\mathbf{P}(t)$. Sea $h : \mathbb{R}^{2^n} \rightarrow \mathbb{R}$ una función vectorial que suma los valores de los componentes:

$$h(\mathbf{x}) = \sum_j x_j \quad (6.17)$$

de tal manera que $\mathbf{P}(t) \in \Lambda$ implica que $h(\mathbf{P}(t)) = 1$. Entonces

$$\mathbf{f}^T\mathbf{P}(t) = h(\text{diag}(\mathbf{f})\mathbf{P}(t)) \quad (6.18)$$

Lema 6.3.1 Sea $\bar{\mathbf{A}}$ una matriz cuyas columnas suman 1. Esto es $\sum_i A_{i,j} = 1$ para cada j . Entonces $h(\bar{\mathbf{A}}\mathbf{x}) = h(\mathbf{x})$ para todo \mathbf{x}

Demostración

$$h(\bar{\mathbf{A}}\mathbf{x}) = \sum_i (\bar{\mathbf{A}}\mathbf{x})_i = \sum_i \sum_j A_{i,j} x_j = \sum_j x_j \sum_i A_{i,j} = \sum_j x_j = h(\mathbf{x}) \quad (6.19)$$

Ahora, la matriz de mutación $\bar{\mathbf{A}}$ tiene la propiedad de que sus columnas suman uno, al igual que la matriz $\bar{\mathbf{Q}}$ si hemos escalado los vectores propios apropiadamente. Por lo tanto

$$\begin{aligned} \mathbf{f}^T \mathbf{P}(t) &= h(\text{diag}(\mathbf{f})\mathbf{P}(t)) \\ &= h(\bar{\mathbf{A}}\text{diag}(\mathbf{f})\mathbf{P}(t)) \\ &= h(\bar{\mathbf{Q}}\text{diag}(\lambda)\tilde{\bar{\mathbf{Q}}}\mathbf{P}(t)) \\ &= h(\text{diag}(\lambda)\tilde{\bar{\mathbf{Q}}}\mathbf{P}(t)) \\ &= \lambda^T \tilde{\bar{\mathbf{Q}}}\mathbf{P}(t) \end{aligned}$$

Poniendo estas ecuaciones juntas, supongamos que el vector de proporciones es $\mathbf{P}(t)$ al tiempo t , entonces

$$\mathbf{P}(t+1) = \mathcal{G}(\mathbf{P}(t)) \quad (6.20)$$

donde \mathcal{G} es el operador generacional que en este caso representa la composición de selección y mutación. Ahora sea $\boldsymbol{\alpha}(t) = \tilde{\bar{\mathbf{Q}}}\mathbf{P}(t)$ la misma población expresada en la nueva base. Entonces tenemos

$$\begin{aligned} \boldsymbol{\alpha}(t+1) &= \tilde{\bar{\mathbf{Q}}}\mathcal{G}(\mathbf{P}(t)) \\ &= \frac{\text{diag}(\lambda)\tilde{\bar{\mathbf{Q}}}\mathbf{P}(t)}{\mathbf{f}^T \mathbf{P}(t)} \\ &= \frac{\text{diag}(\lambda)\tilde{\bar{\mathbf{Q}}}\mathbf{P}(t)}{\lambda^T \tilde{\bar{\mathbf{Q}}}\mathbf{P}(t)} \\ &= \frac{\text{diag}(\lambda)\boldsymbol{\alpha}(t)}{\lambda^T \boldsymbol{\alpha}(t)} \end{aligned}$$

Ahora esta ecuación es idéntica a la derivada para selección proporcional solamente. La única diferencia es que ahora tenemos un vector de "aptitud" λ que contiene los valores propios de $\bar{\mathbf{A}}\text{diag}(\mathbf{f})$. Esto significa que la dinámica del AG con población infinita bajo el operador de mutación y selección proporcional con vector de aptitud \mathbf{f} es idéntica, en la nueva base, a la de dinámica de solo selección proporcional. Entonces podemos aplicar el mismo resultado para obtener las soluciones.

$$\alpha(t)_i = \frac{\lambda_i^t \alpha(0)_i}{\sum_j \lambda_j^t \alpha(t)_j} \quad (6.21)$$

recordando que $\alpha(t) = \widetilde{\mathbf{Q}}\mathbf{P}(t)$ representa la población en la base de vectores propios.

6.4. Solución para la ecuación de cruce

Para ilustrar los principios generales consideraremos nuevamente el problema para 3 bits. La estructura general de las ecuaciones en (4.33) son tales que una solución se construye vía bloques constructores intermedios. Los bloques más fundamentales son los 1-esquemas dado que no pueden ser cortados y por lo tanto su transformación bajo recombinación es trivial (salvo en el caso de cruce uniforme). Explícitamente $P(i**t) = P(i**,0)$, $P(*j*t) = P(*j*,0)$ y $P(**k,t) = P(**k,0)$. Hay cuatro 2-esquemas por punto de cruce correspondientes a $P(ij*t)$ y $P(*jk,t)$, $i, j = 0, 1$. $P(ij*t)$ satisfice

$$P(ij*t) = \exp\left(-\frac{p_c t}{2}\right) P(ij*,0) + \frac{p_c}{2} \exp\left(-\frac{p_c t}{2}\right) \int_0^t P(i**t') P(*j*t') \exp\left(\frac{p_c t'}{2}\right) dt' \quad (6.22)$$

con una ecuación análoga para $P(*jk,t)$. La solución simple para (6.22) es

$$P(ij*t) = \exp\left(-\frac{p_c t}{2}\right) P(ij*,0) + \left(1 - \exp\left(-\frac{p_c t}{2}\right)\right) P(i**t) P(*j*t) \quad (6.23)$$

La solución para 3-esquemas se encuentra usando (6.23)

$$P(ijk,t) = \exp(-p_c t) P(ijk,0) + \frac{p_c}{2} \exp(-p_c t) \int_0^t \exp(p_c t') (P(ij*t') P(**k,t') + (P(i**t') P(*jk,t'))) dt' \quad (6.24)$$

Sustituyendo (6.22) en (6.23) encontramos

$$P(ijk,t) = \exp(-p_c t) P(ijk,0) + \exp\left(-\frac{p_c t}{2}\right) \left(1 - \exp\left(-\frac{p_c t}{2}\right)\right) (P(ij*t) P(**k,t) + (P(i**t) P(*jk,t))) + \left(1 - \exp\left(-\frac{p_c t}{2}\right)\right)^2 P(i**t) P(*j*t) P(**k,t) \quad (6.25)$$

En el límite $t \rightarrow \infty$, $P(ijk,t) \rightarrow P(i**,0) P(*j*,0) P(**k,0)$. Notamos que la aproximación a las proporciones de Robbins es exponencialmente rápida con los esquemas más grandes decayendo más rápido.

La solución general para una cadena de N bits es, como se muestra en [23]

$$P(C_i, t) = \sum_{n=0}^{N-1} \exp\left(-\frac{np_c t}{N-1}\right) \left(1 - \exp\left(-\frac{p_c t}{N-1}\right)\right)^{N-n-1} \mathcal{P}(n+1) \quad (6.26)$$

Donde $\mathcal{P}(n+1)$ es una condición inicial y representa una partición sobre la probabilidad para encontrar $N-n$ bloques constructores al tiempo $t=0$. Para un n dado hay ${}^{N-1}C_n$ de estos términos. Como ya se ha mencionado en diferentes ocasiones, esta solución constituye una prueba para el teorema de Geiringer.

6.5. Comportamiento de la aptitud como función del tiempo considerando solo selección

Siguiendo estas ideas, observemos el comportamiento de la aptitud promedio $\bar{f}(t)$

Por definición:

$$\bar{f}(t) = \sum_{C_i} f(C_i) P(C_i, t)$$

si sustituimos $P(C_i, t)$ por su valor en (6.10) tenemos:

$$\bar{f}(t) = \frac{\sum_{C_i} f^{t+1}(C_i) P(C_i, 0)}{\sum_{C_i} f^t(C_i) P(C_i, 0)} \quad (6.27)$$

Se asume una población inicial aleatoria. Como casos especiales podemos considerar: El escenario de contando unos, en el que encontramos una simetría entre las cadenas con el mismo número de unos. Si N es la longitud de las cadenas, entonces hay ${}^N C_n$ cadenas con n unos, entonces podemos reescribir la suma como:

$$\bar{f}(t) = \frac{\sum_{n=0}^N {}^N C_n n^{t+1}}{\sum_{n=0}^N {}^N C_n n^t} \quad (6.28)$$

Claramente este cociente es una función monótona del tiempo, es decir que la aptitud promedio crece en cada generación

Ahora para el caso del Modelo de Eigen:

$$\bar{f}(t) = \frac{\frac{1}{2^N} f_m^t + (1 - \frac{1}{2^N}) f_o^t}{\frac{1}{2^N} f_m^t + (1 - \frac{1}{2^N}) f_o^t} \quad (6.29)$$

Donde f_m corresponde a la aptitud de la secuencia maestra y f_o a la aptitud de todos los demás miembros de la población. Como podríamos esperar, si trabajamos con cadenas largas ($N \rightarrow \infty$) entonces $\bar{f}(t) = f_o$. Por otro lado, entre más pequeñas las cadenas es mayor la probabilidad de encontrar la secuencia maestra $\bar{f}(t) = f_m$.

Ahora considerando los límites en el tiempo, para pequeños periodos de tiempo ($t \rightarrow 0$) $\bar{f}(t) = f_o$, y para tiempos considerablemente grandes ($t \rightarrow \infty$) podemos alcanzar la secuencia maestra $\bar{f}(t) = f_m$.

Nuevamente, de la ecuación (6.10) asumiendo una población inicial aleatoriamente distribuida tenemos:

$$P(C_i, t) = \frac{f^t(C_i)}{\sum_{C_j} f^t(C_j)} \quad (6.30)$$

Ahora contando unos observamos:

$$P(C_i \in \{C_j | C_j \text{ has } n \text{ ones}\}, t) = \frac{N C_n n^t}{\sum_{n=0}^N N C_n n^t} \quad (6.31)$$

Y en el caso en el que definimos la aptitud como:

$$f(C_i) = \sum_j \sigma_j \sigma_{j+1} \quad (6.32)$$

El número de maneras diferentes de seleccionar k kinks de N bits es $2^{(N-1)C_k}$, entonces de k la aptitud de una cadena dada es $N + k$ y la proporción de una cadena fija es:

$$P(C_i \in \{C_j | C_j \text{ has } k \text{ kinks}\}, t) = \frac{2^{(N-1)C_k} (N + k)}{\sum_{k'=0}^{N-1} 2^{(N-1)C_{k'}} (N + k')} \quad (6.33)$$

6.6. Análisis de valores propios y vectores propios para selección y mutación

Empecemos por hacer una revisión más cercana a la ecuación (5.75). Multiplicamos ambos lados por \bar{Q} para ver como se comportan los vectores propios más claramente

$$\begin{aligned} P(00, t) + P(01, t) + P(10, t) + P(11, t) &= P(00, 0) + P(01, 0) + P(10, 0) + P(11, 0) \\ P(11, t) - P(00, t) &= (1 - 2p)^t (P(11, 0) - P(00, 0)) \\ P(10, t) - P(01, t) &= (1 - 2p)^t (P(10, 0) - P(01, 0)) \\ P(00, t) - P(01, t) - P(10, t) + P(11, t) &= (1 - 2p)^{2t} (P(00, 0) - P(01, 0) - P(10, 0) + P(11, 0)) \end{aligned} \quad (6.34)$$

La primera ecuación claramente muestra la conservación de probabilidades a través del tiempo, en la segunda y tercera vemos como es que las proporciones de las cadenas se acercan al punto de equilibrio exponencialmente rápido, y la última nos muestra la reunión de las dos anteriores, mezclando todo el material genético al doble de velocidad.

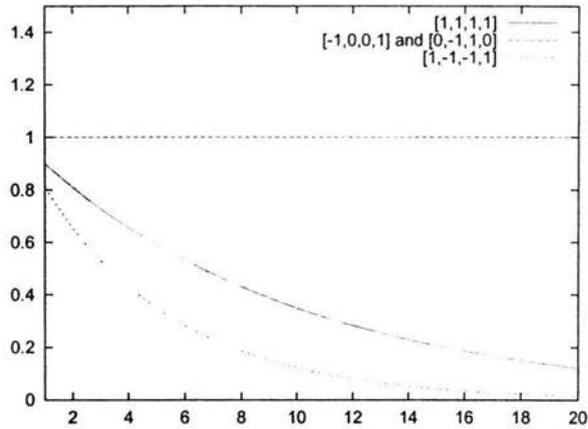


Figura 6.3: Vectores propios como función del tiempo

En la figura 6.3 podemos observar como las proporciones de las cadenas convergen al equilibrio a diferentes velocidades.

Aquí tenemos una tabla de los valores propios con sus vectores propios asociados:

Del cuadro 6.1 podemos inferir la relación entre los valores propios y sus multiplicidades como función de el número de bits. Explícitamente: para N bits hay N valores característicos, todos potencias desde cero hasta N de $1 - 2p$, y la potencia i -ésima tiene multiplicidad ${}^N C_i$

#bits	Eigenvalor	Multipl	Vectores propios
1	1	1	[1, 1]
	$1 - 2p$	1	[-1, 1]
2	1	1	[1, 1, 1, 1]
	$1 - 2p$	2	[0,-1, 1, 0] [-1, 0, 0, 1]
	$(1 - 2p)^2$	1	[-1, 1, 1,-1]
3	1	1	[1, 1, 1, 1, 1, 1, 1, 1]
	$1 - 2p$	3	[-1, 0,-1, 0, 0, 1, 0, 1] [-1, 0, 0, 1,-1, 0, 0, 1] [1, 1, 0, 0, 0, 0,-1,-1]
	$(1 - 2p)^2$	3	[-1, 0, 0, 1, 1, 0, 0,-1] [-1, 1, 0, 0, 0, 0, 1,-1] [-1, 0, 1, 0, 0, 1, 0,-1]
	$(1 - 2p)^3$	1	[1,-1,-1, 1,-1, 1, 1,-1]
4	1	1	[1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1]
	$1 - 2p$	4	[1, 0, 1, 0, 0,-1, 0,-1, 1, 0, 1, 0, 0,-1, 0,-1] [-1, 0, 0, 1,-1, 0, 0, 1,-1, 0, 0, 1,-1, 0, 0, 1] [0, 0, 0, 0,-1,-1,-1,-1, 1, 1, 1, 1, 0, 0, 0, 0] [1, 1, 0, 0, 1, 1, 0, 0, 0, 0,-1,-1, 0, 0,-1,-1]
	$(1 - 2p)^2$	6	[0, 0, 0, 1, 0, 0,-1, 0, 0,-1, 0, 0, 1, 0, 0, 0] [0, 0, 0, 0, 1, 0, 0,-1,-1, 0, 0, 1, 0, 0, 0, 0] [0, 0, 0, 0, 0, 1,-1, 0, 0,-1, 1, 0, 0, 0, 0, 0] [0, 1, 0, 0, 0, 0, 0,-1,-1, 0, 0, 0, 0, 0, 1, 0] [1, 0, 0, 0, 0, 0,-1, 0, 0,-1, 0, 0, 0, 0, 0, 1] [0, 0, 1, 0, 0, 0, 0,-1,-1, 0, 0, 0, 0, 1, 0, 0]
	$(1 - 2p)^3$	4	[1,-1,-1, 1, 0, 0, 0, 0, 0, 0,-1, 1, 1,-1] [-1, 0, 0, 1, 1, 0, 0,-1, 1, 0, 0,-1,-1, 0, 0, 1] [1,-1, 0, 0,-1, 1, 0, 0, 0, 0,-1, 1, 0, 0, 1,-1] [0,-1, 0, 1, 1, 0,-1, 0, 0, 1, 0,-1,-1, 0, 1, 0]
$(1 - 2p)^4$	1	[-1, 1, 1,-1, 1,-1,-1, 1, 1,-1,-1, 1,-1, 1, 1,-1]	

Cuadro 6.1: Tabla de Vectores propios

Capítulo 7

Conclusiones

A pesar de todo el trabajo desarrollado hasta el momento en teoría de AGs, todavía estamos lejos de poder dar “reglas del pulgar” para ajustar precisamente los parámetros tales como el tamaño de la población o la probabilidad de mutación con un fundamento teórico. Los parámetros, por ahora, se ajustan siguiendo reglas obtenidas empíricamente y ajustándolos al escenario mediante prueba y error.

El presente trabajo nos ha servido para obtener intuición sobre el comportamiento de los AGs en algunos escenarios que consideramos de importancia, mostramos como estos escenarios nos pueden dar una perspectiva local de un AG actuando en un espacio de búsqueda, vimos como estas localizaciones pueden ser en espacio, considerando clases de problemas con función de aptitud similar, o en tiempo, como cuando buscamos en las etapas tempranas o cercanas a la convergencia del AG.

Desarrollamos el formalismo de las ecuaciones exactas planteadas por Stephens [23, 24, 25, 26, 27, 28] tanto a nivel de cadenas como con el cambio de escala de esquemas, mostrando como los esquemas son un cambio de escala que surge naturalmente a partir de la cruce. Siguiendo esta idea, usamos cambios de escala para simplificar las ecuaciones en los escenarios que planteamos. De esto concluimos que el cambio de escala es un instrumento que nos permite sacar provecho de las simetrías que inducen los operadores genéticos.

Mostramos las soluciones analíticas a sistemas con pocos grados de libertad y en algunos casos mostramos generalizaciones naturales a sistemas más grandes. Usamos el concepto de cambio de escala para encontrar los grados de libertad efectivos con ejemplos específicos, probando nuevamente la trascendencia del cambio de escala.

Observamos el comportamiento de las soluciones en el simplex y encontramos puntos fijos.

Recopilamos algunas soluciones a las ecuaciones de evolución que nos permiten observar comportamientos en escenarios más generales a los que planteamos. De esta manera mostramos todo un compendio de resultados, desde los más simples hasta los más generales. En algunos casos la extensión de los casos simples

a los generales es bastante natural pero en otros hace falta más trabajo y los resultados no siempre son intuitivos.

Condujimos una serie de experimentos con los que podemos constatar la validez de las ecuaciones. Para un conjunto de parámetros repetimos el experimento muchas veces y con eso obtuvimos el comportamiento promedio que después comparamos con los resultados de la iteración directa de las ecuaciones. Observamos que a medida que se incrementa el número de experimentos a promediar, la curva obtenida se hace más suave y parecida a la de la iteración. Con esto ahora tenemos una medida intuitiva de los efectos de tamaño finito.

7.1. Reconocimientos

Me gustaría agradecer al Dr. Jon Rowe y al prof. Xin Yao por la ayuda para el intercambio académico parcialmente patrocinado por la DGEP-UNAM dentro del programa de movilidad internacional de estudiantes. Este trabajo de tesis fue apoyado parcialmente por la DGAPA proyecto IN100201.

Apéndice A

Soluciones alternativas y aproximaciones tomando límite de tiempo continuo

A.1. Soluciones para selección y mutación 1 bit escenario de contando 1's

Si en la ecuación (5.50) asignamos aptitud α a la cadena 0 y $1 + \alpha$ a 1 tenemos:

$$\begin{pmatrix} P(0, t+1) \\ P(1, t+1) \end{pmatrix} = \begin{pmatrix} \frac{\alpha(1-p_m)}{\alpha P(0,t)+(1+\alpha)P(1,t)} & \frac{(1+\alpha)p_m}{\alpha P(0,t)+(1+\alpha)P(1,t)} \\ \frac{\alpha p_m}{\alpha P(0,t)+(1+\alpha)P(1,t)} & \frac{(1+\alpha)(1-p_m)}{\alpha P(0,t)+(1+\alpha)P(1,t)} \end{pmatrix} \begin{pmatrix} P(0, t) \\ P(1, t) \end{pmatrix} \quad (\text{A.1})$$

Equivalentemente

$$\begin{pmatrix} P(0, t+1) \\ P(1, t+1) \end{pmatrix} = \begin{pmatrix} \frac{\alpha(1-p_m)}{\alpha+P(1,t)} & \frac{(1+\alpha)p_m}{\alpha+P(1,t)} \\ \frac{\alpha p_m}{\alpha+P(1,t)} & \frac{(1+\alpha)(1-p_m)}{\alpha+P(1,t)} \end{pmatrix} \begin{pmatrix} P(0, t) \\ P(1, t) \end{pmatrix} \quad (\text{A.2})$$

Si sustituimos $P(1, t)$ por $1 - P(0, t)$ en la primera ecuación y $P(0, t)$ por $1 - P(1, t)$ en la segunda y simplificamos tenemos las siguientes ecuaciones:

$$P(0, t+1) = \frac{p_m(1+\alpha) + [\alpha(1-p_m) - p_m(1+\alpha)]P(0, t)}{1 + \alpha + P(0, t)} \quad (\text{A.3})$$

$$P(1, t+1) = \frac{\alpha p_m + [(1+\alpha)(1-p_m) - \alpha p_m]P(1, t)}{\alpha + P(1, t)} \quad (\text{A.4})$$

Ambas tienen la forma $y(t+1) = \frac{a+by(t)}{c+ey(t)}$. Si restamos $y(t)$ de ambos lados y simplificamos tenemos:

$$\frac{y(t+1) - y(t)}{1} = \frac{a + b'y(t) + fy(t)}{c + ey(t)} \quad (\text{A.5})$$

donde $b' = b - c$ y $f = -e$. Ahora si tomamos el límite de tiempo continuo observamos

$$\frac{dy(t)}{dt} = \frac{a + b'y(t) + fy(t)}{c + ey(t)} \quad (\text{A.6})$$

y despejando dt obtenemos la ecuación:

$$\int dt = \int \frac{c + cy(t)}{a + b'y(t) + fy(t)} dy \quad (\text{A.7})$$

Integrando y tomando límite finalmente tenemos:

$$t = \tanh^{-1} \left(\frac{1 - p_m - 2y}{1 - p_m} \right) - \frac{\ln((1 - p_m)y - y^2)}{2} \quad (\text{A.8})$$

Y para $P(0, t)$ tenemos:

$$t = \tanh^{-1} \left(\frac{-p_m - 1 + 2y}{\sqrt{4p_m + (-p_m - 1)^2}} \right) \frac{(1 - p_m)}{\sqrt{4p_m + (-p_m - 1)^2}} - \frac{\ln(-p_m + (-p_m - 1)y + y^2)}{2} \quad (\text{A.9})$$

Estas ecuaciones nos muestran el comportamiento para el límite cuando la selección es muy débil alrededor del punto inicial. Si tomamos el límite $\alpha \rightarrow 0$ desde un inicio obtenemos:

$$\begin{pmatrix} P(0, t+1) \\ P(1, t+1) \end{pmatrix} = \begin{pmatrix} 0 & \frac{p_m}{P(1, t)} \\ 0 & \frac{(1-p_m)}{P(1, t)} \end{pmatrix} \begin{pmatrix} P(0, t) \\ P(1, t) \end{pmatrix} \quad (\text{A.10})$$

cuyas soluciones son:

$$\lim_{t \rightarrow \infty} \begin{pmatrix} P(0, t) \\ P(1, t) \end{pmatrix} = \begin{pmatrix} p_m \\ 1 - p_m \end{pmatrix} \quad (\text{A.11})$$

para $t > 0$. Este resultado nos indica la existencia del punto fijo al que debe aproximarse la solución.

A.2. Soluciones para selección y mutación 1 bit modelo de Eigen

$$\begin{pmatrix} P(0, t+1) \\ P(1, t+1) \end{pmatrix} = \begin{pmatrix} 1 - p_m & p_m \\ p_m & 1 - p_m \end{pmatrix} \begin{pmatrix} \frac{f_0}{f_0 P(0, t) + f_{sm} P(1, t)} & 0 \\ 0 & \frac{f_{sm}}{f_0 P(0, t) + f_{sm} P(1, t)} \end{pmatrix} \begin{pmatrix} P(0, t) \\ P(1, t) \end{pmatrix} \quad (\text{A.12})$$

Si sustituimos $P(0, t)$ por $1 - P(1, t)$ y simplificamos tenemos las siguientes ecuaciones:

$$P(0, t+1) = \frac{(1 - p_m)f_0 + (p_m(f_{sm} - f_0) + f_0)P(1, t)}{f_0 + (f_{sm} + f_0)P(1, t)} \quad (\text{A.13})$$

$$P(1, t+1) = \frac{(1 - p_m)f_0 + (p_m(f_0 - f_{sm}) + f_{sm})P(1, t)}{f_0 + (f_{sm} + f_0)P(1, t)} \quad (\text{A.14})$$

De aquí:

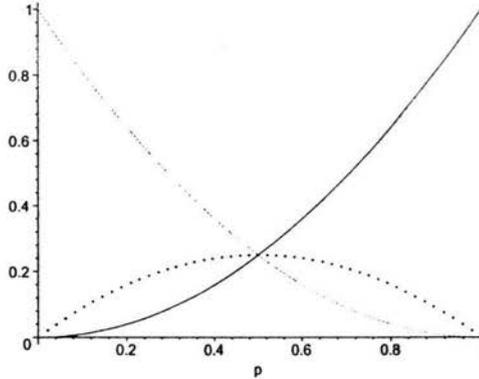


Figura A.1: Selección y mutación para 2 bits Modelo de Eigen: Comportamiento asintótico de $P(00, t) = p_m^2$ (línea continua) $P(01, t) = P(10, t) = p_m(1 - p_m)$ (cruces) y $P(11, t) = (1 - p_m)^2$ (línea discontinua)

Nuevamente 2 bits son pocos para poder contemplar las características del modelo pero esta representación matricial nos deja ver con facilidad que la entrada que contiene a f_{sm} con las sucesivas iteraciones crecerá exponencialmente. De hecho con ejemplos numéricos tomando $f_0 \sim f_{sm}/10$ nos percatamos que en solo un par de iteraciones la proporción de la secuencia maestra sobrepasa el 99% para $p_m = 0.05$.

A.4. Soluciones para selección y cruza 2 bits escenario plano

$$\begin{aligned} P(00, t + 1) &= P(00, t) - p_c[\Delta(t)] \\ P(01, t + 1) &= P(01, t) + p_c[\Delta(t)] \\ P(10, t + 1) &= P(10, t) + p_c[\Delta(t)] \\ P(11, t + 1) &= P(11, t) - p_c[\Delta(t)] \end{aligned}$$

Donde $\Delta(t) = P(00, t)P(11, t) - P(01, t)P(10, t)$ es el "Linkage Disequilibrium Coefficient". Notemos que se llega a un punto de equilibrio precisamente cuando $\Delta(t)$ es cero. Dado que no hay presión selectiva el efecto de la cruza es de homogeneización de las proporciones.

Consideremos $\alpha(t) = [P(00, t) + P(11, t)] - [P(01, t) + P(10, t)]$ y $\Delta(t)$ como antes. Si escribimos $\Delta(t + 1)$ y $\alpha(t + 1)$ en términos de $\Delta(t)$ y $\alpha(t)$ obtenemos

Ambas tienen la forma $y(t+1) = \frac{a+by(t)}{c+ey(t)}$. Si restamos $y(t)$ de ambos lados y simplificamos tenemos:

$$\frac{y(t+1) - y(t)}{1} = \frac{a + b'y(t) + fy(t)}{c + ey(t)} \quad (\text{A.15})$$

donde $b' = b - c$ y $f = -e$. Ahora si tomamos el límite de tiempo continuo observamos

$$\frac{dy(t)}{dt} = \frac{a + b'y(t) + fy(t)}{c + ey(t)} \quad (\text{A.16})$$

y despejando dt obtenemos la ecuación:

$$\int dt = \int \frac{c + ey(t)}{a + b'y(t) + fy(t)} dy \quad (\text{A.17})$$

$$t = \frac{((1 - p_m)(f_{sm} - f_0) + 2f_0) \tanh^{-1} \left(\frac{(1-p_m)(f_{sm}-f_0) - 2(f_{sm}+f_0)P(1,t)}{\sqrt{4p_m f_0 (f_{sm}+f_0) + (1-p_m)^2 (f_{sm}-f_0)^2}} \right)}{\sqrt{4p_m f_0 (f_{sm} + f_0) + (1 - p_m)^2 (f_{sm} - f_0)^2}} - \frac{\ln \left(p_m f_0 + (1 - p_m)(f_{sm} - f_0) P(1, t) - (f_{sm} + f_0) P(1, t)^2 \right)}{2} \quad (\text{A.18})$$

Y para $P(0, t)$ tenemos:

$$t = \frac{(p_m (f_{sm} - f_0) + 2f_0) \tanh^{-1} \left(\frac{(f_{sm}(p_m-2) - f_0(p_m+2)) - 2(f_{sm}+f_0)P(0,t)}{\sqrt{4(1-p_m)f_0(f_{sm}+f_0) + p_m^2(f_{sm}-f_0)^2}} \right)}{\sqrt{4(1-p_m)f_0(f_{sm} + f_0) + p_m^2(f_{sm} - f_0)^2}} - \frac{\ln \left(f_0(1 - 3p_m) - f_{sm}(1 - p_m) + (f_{sm}(p_m - 2) - f_0(p_m + 2)) P(0, t) - (f_{sm} + f_0) P(0, t)^2 \right)}{2} \quad (\text{A.19})$$

Notemos que esta es la solución general para selección y mutación para 1 bit, dado que no hemos hecho suposiciones sobre f_0 o f_{sm} .

A.3. Soluciones para selección y mutación 2 bits modelo de Eigen

Suponiendo $f_{sm} \gg f_0$, normalizamos por f_{sm} y tenemos:

$$\begin{pmatrix} P(00, t+1) \\ P(01, t+1) \\ P(10, t+1) \\ P(11, t+1) \end{pmatrix} = \frac{1}{P(11, t)} \begin{pmatrix} 0 & 0 & 0 & p^2 \\ 0 & 0 & 0 & (1-p)p \\ 0 & 0 & 0 & (1-p)p \\ 0 & 0 & 0 & (1-p)^2 \end{pmatrix} \begin{pmatrix} P(00, t) \\ P(01, t) \\ P(10, t) \\ P(11, t) \end{pmatrix} \quad (\text{A.20})$$

el siguiente sistema de ecuaciones:

$$\alpha(t+1) = \alpha(t) - 4p_c\Delta(t) \quad (\text{A.21})$$

$$\Delta(t+1) = \Delta(t) - p_c\alpha(t)\Delta(t) \quad (\text{A.22})$$

Pasando $\alpha(t)$ (respectivamente $\Delta(t)$) a la izquierda y tomando el límite de tiempo continuo tenemos:

$$\frac{d}{dt}\alpha(t) = -4p_c\Delta(t) \quad (\text{A.23})$$

$$\frac{d}{dt}\Delta(t) = -p_c\alpha(t)\Delta(t) \quad (\text{A.24})$$

cuyas soluciones son:

$$\alpha(t) = \frac{\tanh(1/2\sqrt{C_1 p_c}\sqrt{2}(t+C_2))\sqrt{C_1 p_c}\sqrt{2}}{p_c} \quad (\text{A.25})$$

$$\Delta(t) = -1/4 \frac{(1 - (\tanh(1/2\sqrt{C_1 p_c}\sqrt{2}(t+C_2)))^2) C_1}{p_c} \quad (\text{A.26})$$

Tomando como condiciones iniciales $\Delta(0) = P(00,0)P(11,0) - P(01,0)P(10,0)$ y $\alpha(0) = [P(00,0) + P(11,0)] - [P(01,0) + P(10,0)]$ obtenemos los valores para las constantes C_1 y C_2 :

$$C_1 = \frac{\alpha(0)^2 p_c}{2} - 4\Delta(0)p_c \quad (\text{A.27})$$

$$C_2 = \frac{\sqrt{2}}{p_c\sqrt{\frac{\alpha(0)^2}{2} - 4\Delta(0)}} \tanh^{-1}\left(\frac{\alpha(0)}{\sqrt{\frac{\alpha(0)^2}{2} - 4\Delta(0)}}\right) \quad (\text{A.28})$$

De las ecuaciones A.23 y A.24 observamos que, de existir un punto fijo, este se tiene cuando: $P(00,t)P(11,t) = P(01,t)P(10,t)$. Si agregamos la ecuación de conservación de probabilidad $P(00,t) + P(11,t) + P(01,t) + P(10,t) = 1$, podemos concluir inmediatamente (despejando $P(00,t)$ de la segunda ecuación y sustituyendo en la primera) que $P(00,t) + P(11,t) = \frac{1}{2} = P(01,t) + P(10,t)$. La existencia del punto fijo la garantiza la ecuación A.26 pues claramente $\lim_{t \rightarrow \infty} \Delta(t) = 0$; de hecho hay una infinidad de puntos fijos que dependen de las condiciones iniciales. Ahora, si asumimos una distribución aleatoria en las condiciones iniciales, el punto fijo para cada proporción es: $P(00,t) = P(11,t) = P(01,t) = P(10,t) = \frac{1}{4}$. En otras palabras, este punto es un invariante bajo el operador de cruce en un escenario plano.

A.5. Soluciones para selección y cruce 3 bits escenario plano

Como podemos observar en el caso de 2 bits, las ecuaciones que representan la evolución de las cadenas al aplicarles el operador de cruce son complicadas.

En esta sección haremos uso del cambio de escala de esquemas para mostrar los beneficios del cambio de base. Consideremos la cuerda 111, la siguiente ecuación muestra la matriz de cambio a la base de bloques constructores de 111:

$$\begin{pmatrix} P(111, t) \\ P(11*, t) \\ P(1*1, t) \\ P(1**, t) \\ P(*11, t) \\ P(*1*, t) \\ P(**1, t) \\ P(***, t) \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \end{pmatrix} \begin{pmatrix} P(111, t) \\ P(110, t) \\ P(101, t) \\ P(100, t) \\ P(011, t) \\ P(010, t) \\ P(001, t) \\ P(000, t) \end{pmatrix} \quad (\text{A.29})$$

Entonces la ecuación para la proporción de 111 después de la cruza queda como:

$$P(111, t+1) = (1 - p_c)P(111, t) + p_c [P(1**, t)P(*11, t) + P(11*, t)P(**1, t)] \quad (\text{A.30})$$

En esta nueva base podemos escribir entonces

$$P(*11, t) = (1 - p_c)P(*11, t-1) + \frac{p_c}{2} [P(*1*, t-1)P(**1, t-1)] \quad (\text{A.31})$$

$$P(11*, t) = (1 - p_c)P(11*, t-1) + \frac{p_c}{2} [P(*1*, t-1)P(1**, t-1)] \quad (\text{A.32})$$

Podemos resolver estas ecuaciones transformándolas a ecuaciones diferenciales de la siguiente manera:

$$\frac{P(*11, t) - P(*11, t-1)}{1} = p_c P(*11, t-1) + \frac{p_c}{2} [P(*1*, t-1)P(**1, t-1)] \quad (\text{A.33})$$

$$\frac{P(11*, t) - P(11*, t-1)}{1} = p_c P(11*, t-1) + \frac{p_c}{2} [P(*1*, t-1)P(1**, t-1)] \quad (\text{A.34})$$

Cuyas soluciones son:

$$\begin{aligned} P(*11, t) &= \exp\left(-\frac{p_c t}{2}\right) P(*11, 0) + \\ &\frac{p_c}{2} \exp\left(-\frac{p_c t}{2}\right) \int_0^t P(*1*, t') P(**1, t') \exp\left(-\frac{p_c t'}{2}\right) dt' \\ P(11*, t) &= \exp\left(-\frac{p_c t}{2}\right) P(11*, 0) + \\ &\frac{p_c}{2} \exp\left(-\frac{p_c t}{2}\right) \int_0^t P(*1*, t') P(1**, t') \exp\left(-\frac{p_c t'}{2}\right) dt' \end{aligned} \quad (\text{A.35})$$

Unas soluciones inmediatas para estas ecuaciones son:

$$P(*11, t) = \exp\left(-\frac{p_c t}{2}\right) P(*11, 0) + \left(1 - \exp\left(-\frac{p_c t}{2}\right)\right) P(*1*, 0) P(**1, 0) \quad (\text{A.36})$$

$$P(11*, t) = \exp\left(-\frac{p_c t}{2}\right) P(11*, 0) + \left(1 - \exp\left(-\frac{p_c t}{2}\right)\right) P(*1*, 0) P(1**, 0) \quad (\text{A.37})$$

Uniendo estas soluciones tenemos la solución para A.30 que es la siguiente:

$$\begin{aligned}
P(111, t) = & \exp(-p_c t) P(111, 0) + \exp\left(-\frac{p_c t}{2}\right) \left(1 - \exp\left(-\frac{p_c t}{2}\right)\right) \\
& \times (P(11*, 0)P(**1, 0) + P(*11, 0)P(1**, 0)) \quad (\text{A.38}) \\
& + \left(1 - \exp\left(-\frac{p_c t}{2}\right)\right)^2 P(1**, 0)P(*1*, 0)P(**1, 0)
\end{aligned}$$

Notemos que no se ha tomado ninguna característica especial de 111 así que, si la sustituimos por la cadena ijk , $i, j, k \in [0, 1]$ el resultado permanece sin alteraciones, es decir:

$$\begin{aligned}
P(ijk, t) = & \exp(-p_c t) P(ijk, 0) + \exp\left(-\frac{p_c t}{2}\right) \left(1 - \exp\left(-\frac{p_c t}{2}\right)\right) \\
& \times (P(ij*, 0)P(**k, 0) + P(*jk, 0)P(i**, 0)) + \quad (\text{A.39}) \\
& \left(1 - \exp\left(-\frac{p_c t}{2}\right)\right)^2 P(i**, 0)P(*j*, 0)P(**k, 0)
\end{aligned}$$

En el límite $t \rightarrow \infty$ tenemos $P(ijk, t) \rightarrow P(i**, 0)P(*j*, 0)P(**k, 0)$

Que al igual que en el caso discreto coincide con el teorema de Geiringer [8] para tres bits.

A.6. Derivación para la ecuación de mutación 3 bits contando 1's

Para 3 bits contando 1's ,denotemos $z(i, j) = p^{d(C_i, C_j)}(1 - p)^{N-d(C_i, C_j)}$ y W_k al conjunto de cadenas con k 1's.

Veamos por ejemplo $P(W_1, t + 1) =$

$$\begin{aligned}
& z(001, 000)P(000, t) + z(010, 000)P(000, t) + z(100, 000)P(000, t) \\
& + z(001, 001)P(001, t) + z(001, 010)P(010, t) + z(001, 100)P(100, t) \\
& + z(010, 001)P(001, t) + z(010, 010)P(010, t) + z(010, 100)P(100, t) \\
& + z(100, 001)P(001, t) + z(100, 010)P(010, t) + z(100, 100)P(100, t) \\
& + z(001, 011)P(011, t) + z(001, 101)P(101, t) + z(001, 110)P(110, t) \\
& + z(010, 011)P(011, t) + z(010, 101)P(101, t) + z(010, 110)P(110, t) \\
& + z(100, 011)P(011, t) + z(100, 101)P(101, t) + z(100, 110)P(110, t) \\
& + z(001, 111)P(111, t) + z(010, 111)P(111, t) + z(100, 111)P(111, t)
\end{aligned} \quad (\text{A.40})$$

ahora denotemos por $z[k] = p^k(1-p)^{N-k}$ y tenemos $P(W_1, t+1) =$

$$\begin{aligned}
& z[1]P(000, t)z[1]P(000, t)z[1]P(000, t) \\
& +z[0]P(001, t) + z[2]P(010, t) + z[2]P(100, t) \\
& +z[2]P(001, t) + z[0]P(010, t) + z[2]P(100, t) \\
& +z[2]P(001, t) + z[2]P(010, t) + z[0]P(100, t) \\
& +z[1]P(011, t) + z[1]P(101, t) + z[3]P(110, t) \\
& +z[1]P(011, t) + z[3]P(101, t) + z[1]P(110, t) \\
& +z[3]P(011, t) + z[1]P(101, t) + z[1]P(110, t) \\
& +z[2]P(111, t) + z[2]P(111, t) + z[2]P(111, t)
\end{aligned} \tag{A.41}$$

Si factorizamos tenemos $P(W_1, t+1) =$

$$\begin{aligned}
& (z[1] + z[1] + z[1])P(W_0, t) \\
& +(z[0] + z[2] + z[2])P(W_1, t) \\
& +(z[1] + z[1] + z[3])P(W_2, t) \\
& +(z[2] + z[2] + z[2])P(W_3, t)
\end{aligned} \tag{A.42}$$

En general tenemos

$$P(W_i, t+1) = \sum_{k=0}^{N-1} \mathcal{P}_1(W_i \rightarrow W_k)P(W_k, t) \tag{A.43}$$

donde

$$\mathcal{P}_1(W_i \rightarrow W_k) = \sum_{C_i \in W_i} p^{d(C_i, C_j)}(1-p)^{N-d(C_i, C_j)} \tag{A.44}$$

con $C_j \in W_k$.

Apéndice B

Coarse-graining in genetic algorithms: some issues and examples

Coarse-graining in genetic algorithms: some issues and examples

Andrés Aguilar Contreras¹, Jonathan E. Rowe², and Christopher R. Stephens³

¹ Instituto de Investigación en Matemáticas Aplicadas y Sistemas, UNAM, Circuito Escolar, Ciudad Universitaria 04510, México D.F. 04510
andresac@uxmcc2.iimas.unam.mx

² School of Computer Science, University of Birmingham, Birmingham B15 2TT, Great Britain
J.E.Rowe@cs.bham.ac.uk

³ Instituto de Ciencias Nucleares, UNAM, Circuito Exterior, A. Postal 70-543, México D.F. 04510.
stephens@nuclecu.unam.mx

Abstract. Following the work of Stephens and coworkers on the coarse-grained dynamics of genetic systems, we work towards a possible generalisation in the context of genetic algorithms, giving as examples schemata, genotype-phenotype mappings, and error classes in the Eigen model. We discuss how the dynamics transforms under a coarse-graining, comparing and contrasting different notions of invariance. We work out some examples in the two-bit case, to illustrate the ideas and issues. We then find a bound for the Selection Weighted Linkage Disequilibrium Coefficient for the two-bit onemax problem.

1 Introduction

To model the exact evolution of a genetic algorithm requires us, in general, to track what happens to each possible individual. For example, if the search space is binary strings of length ℓ , we have evolution equations for each of the 2^ℓ possible strings. It may also be of interest to investigate what happens to certain subsets of individuals. There are three reasons for doing this. Firstly, it may be possible to reduce the number of degrees of freedom in the evolution equations and so make a more tractable model. This is particularly true when modelling the appropriate effective degrees of freedom for the dynamics. Secondly, one may be interested in the evolutionary history of one particular individual, and there may be only a limited number of subsets to which its ancestors could have belonged. Thirdly, of course, one may have some intrinsic interest in a certain subset (for example, it may be a subset of high quality individuals or represent some “kinship” or genetically related group such as the individuals associated with a “niche” or a species).

The idea of tracking subsets is the basis of Holland’s schemata [Holland, 1975], Radcliffe’s forma [Radcliffe, 1992] and Vose’s predicates [Vose, 1991]. More recently, Stephens [Stephens and Waelbroek, 1997] has formally studied evolution equations under this kind of coarse-graining using schemata and extended the analysis to other contexts than evolutionary computation in [Stephens, 2003] and [Stadler and Stephens, 2003]. Van Nimwegen [van Nimwegen et al., 1997] has modelled the dynamics of GAs on

royal-road, and other functions, using approximate coarse-grained models. Rowe [Rowe, 1998] has considered the use of unitation classes as a basis for a coarse-grained model of selection-mutation algorithms. We intend to extend this work by looking at possible generalisations and limitations in the context of genetic algorithms.

In [Stephens and Waelbroek, 1997] it was shown that the dynamical equations governing the evolution of a GA with proportional selection, mutation and one-point crossover was *form invariant* under a coarse graining to schemata. This was later extended [Stephens, 2001] to any selection, mutation and homologous crossover operators. This form invariance was later studied by Vose and Wright [Vose and Wright, 2001] who discussed a more restrictive form of the invariance using the notion of *compatibility* [Vose, 1999] between the coarse graining and the genetic operators.

In the following section, we formally define the idea of coarse-graining. We then consider how the dynamics looks after a coarse graining. The ideas and issues are then illustrated by a series of examples. Finally, we apply coarse-graining to help us estimate what happens to the linkage disequilibrium coefficient in the two-bit onemax problem.

2 Coarse-grained dynamics

Let $\Omega = \{0, 1, 2, \dots, n-1\}$ be the search space. In the case of binary strings of length ℓ , we identify each string with an integer under standard binary encoding, and $n = 2^\ell$. We represent a population by a vector $p = (p_0, p_1, \dots, p_{n-1})$ in which p_k is the proportion of individual k in the population. Population vectors are elements of the *simplex*

$$\Lambda = \left\{ p \in \mathbb{R}^n : \sum_k p_k = 1, \text{ and } p_k \geq 0 \text{ for all } k \right\}$$

A coarse-graining of Ω will be a collection of subsets of Ω . Given a fitness function $f : \Omega \rightarrow \mathbb{R}$ we wish to define the fitness of a given subset. Notice, that this will, in general, depend on the details of the population. One can think of this situation as being analogous to a *co-evolutionary* model, in which the fitness of an individual depends on the current population. Let $\mathcal{P}(\Omega)$ denote the *power set* of Ω (that is, the set of all subsets of Ω). Then, formally, we have a function

$$F : \Lambda \rightarrow (\mathcal{P}(\Omega) \rightarrow \mathbb{R})$$

defined as

$$F(p)(A) = \frac{\sum_{i \in A} p_i f(i)}{\sum_{i \in A} p_i}$$

That is, given a population $p \in \Lambda$, $F(p)$ is a “fitness function” which assigns fitnesses to subsets of Ω . The fitness of a subset A is the average fitness of elements of A in population p .

Definition 1. Let $\Gamma = \{\gamma_i\} \subseteq \mathcal{P}(\Omega)$ be a collection of subsets of the search space that covers the search space. That is,

$$\bigcup \gamma_i = \Omega$$

We call such a collection a coarse-graining of Ω . For any population $p \in \Lambda$ we can assign a fitness to each element of Γ using the function $F(p)$.

A coarse-graining is non-degenerate if Γ is a partition of the search space

$$i \neq j \implies \gamma_i \cap \gamma_j = \emptyset$$

Notice that the fitness of a subset in a coarse-graining depends on the current population and therefore on time, even if the underlying fitness function f is static.

Examples

- 1) **Schemata** We can associate a schema with the set of all strings which match it. The set of all schemata forms a highly degenerate cover of the search space, however given an arbitrary choice of string then all the schemata that contain the string forms a new basis of the same dimensionality as the original - the Building Block Basis [Stephens, 2003]. However, unless the fitness function is a constant for all strings matching a given schema, then the fitness of the schema itself will be a dynamic quantity (that is, it will depend on the details of the current population).
- 2) **Genotype-phenotype mappings** Suppose we have a map $\varphi : \Omega \rightarrow \Phi$ which maps genotypes to phenotypes, where Φ is the space of phenotypes. Fitness is then assessed via an individual's phenotype. That is, there is a function $g : \Phi \rightarrow \mathbb{R}$. The fitness of a genotype is then $f = g \circ \varphi$. We can create a non-degenerate coarse-graining by considering subsets of Ω which map to the same phenotype. That is, for each $i \in \Phi$, set

$$\gamma_i = \{a \in \Omega : \varphi(a) = i\}$$

The fitness of such a subset is constant:

$$\begin{aligned} F(p)(\gamma_i) &= \frac{\sum_{j \in \gamma_i} p_j f(j)}{\sum_{j \in \gamma_i} p_j} \\ &= \frac{\sum_{\varphi(j)=i} p_j g \circ \varphi(j)}{\sum_{j \in \gamma_i} p_j} \\ &= \frac{g(i) \sum_{\varphi(j)=i} p_j}{\sum_{j \in \gamma_i} p_j} \\ &= g(i) \end{aligned}$$

This coarse-graining is natural with respect to selection, as we only need to keep track of what happens to the subsets, without worrying about their detailed composition.

- 2.1 **Unitation** A particular example of a genotype-phenotype mapping is when we have a function of unitation. That is, the search space is binary strings of length ℓ and fitness only depends on the number of ones in a string. The phenotype set is

$$\Phi = \{0, 1, 2, \dots, \ell\}$$

2.2 The Eigen Model (Needle-in-a-haystack) A second example is that of the Eigen model [Eigen, 1971]. In this landscape all the strings have the same fitness except for a special string (the optimum) that has a relatively high fitness. - the so called “master sequence”. In this landscape, the genotype-phenotype coarse-graining creates only two equivalence classes, hence there is a reduction in degrees of freedom from N to one.

3 Exact and Approximate invariance under a coarse-graining

Having motivated the idea of coarse graining and given some simple examples one needs to understand how the evolution equations for the GA look under the coarse graining. As mentioned, it was shown in [Stephens and Waelbroek, 1997] that the canonical GA is form invariant under a coarse graining to schemata, i.e. that the equations have exactly the same functional form after such a coarse graining. This is a highly non-trivial result, as a coarse graining in general will not preserve the functional form, as can be simply seen, for example, in the case of coarse graining from genotype to phenotype in the presence of mutation or crossover. Vose later showed [Vose, 1999] that schemata are the only coarse-grained variable that leave the dynamical equations for homologous crossover invariant and hence form a privileged set. However, Vose also introduced a more restrictive form of invariance under coarse graining - *compatibility* - wherein it was not sufficient that the equations be form invariant. Formally, if an operator $\mathcal{M} : A \rightarrow A$ gives the effect of applying an operator to a population, then a coarse-graining (Γ) is compatible with \mathcal{M} if and only if, for any two populations $x, y \in A$

$$\sum_{j \in \gamma_i} x_j = \sum_{j \in \gamma_i} y_j \implies \sum_{j \in \gamma_i} \mathcal{M}(x)_j = \sum_{j \in \gamma_i} \mathcal{M}(y)_j$$

for all $\gamma_i \in \Gamma$ (see chapter 16-17 of [Vose, 1999]). It is known, for example, that schemata are compatible with crossover (by masks) and that unication classes are compatible with mutation.

A simple example illustrates the difference between the two different notions of invariance and coarse graining. Consider selection only in a two-bit one-max model: The equation of motion for proportional selection is

$$P(h_1 h_2, t + 1) = (f(h_1 h_2) / \bar{f}(t)) P(h_1 h_2, t). \quad (1)$$

We pass to the schema h_1^* by coarse graining h_2 to find

$$P(h_1^*, t + 1) = (f(h_1^*, t) / \bar{f}(t)) P(h_1^*, t) \quad (2)$$

where $f(h_1^*, t) = (f(h_1 h_2) P(h_1 h_2, t) + f(h_1 \bar{h}_2) P(h_1 \bar{h}_2, t)) / (P(h_1 h_2, t) + P(h_1 \bar{h}_2, t))$ and \bar{h}_2 is the bit complement of h_2 . Clearly (1) and (2) have the same functional form. However, to satisfy compatibility $f(h_1^*, t)$ would have to satisfy $f(h_1^*, t) = f(h_1)$, where $f(h_1 = 1) = 1$ and $f(h_1 = 0) = 0$. Compatibility will only be valid when the problem exhibits an exact equivalence relation (“symmetry”) and the genetic operators respect this symmetry, such as is the case for the genotype-phenotype map and selection only, or with schemata and crossover only. The existence of an exact symmetry usually

allows for a reduction in the number of degrees of freedom by going to those effective degrees of freedom that are invariant under the symmetry.

However, the utility of coarse graining is not restricted to when it is compatible with the dynamics. For instance, in [Stephens, 2001] the form invariance of the equations of motion was used to prove a generalization of Geiringer’s theorem to the case of non-flat landscapes. Further, in the physical sciences, where coarse graining has played an essential role, its utility is precisely for those cases where symmetries are not present, but rather where the coarse-grained dynamics can provide an approximate description of the system. In this case the closer are the coarse-grained variables to the true effective degrees of freedom then the better the approximation. For instance, in a strong selection regime one would expect phenotypes to approximate well the true dynamics with mutation and/or crossover inducing a small “interaction” between different phenotypes. Similarly, in the case of strong crossover and weak selection one would expect one-schemata to approximate well the dynamics with selection inducing a small interaction between the different one-schemata.

4 Crossover and schemata coarse graining

We can track the evolutionary history of the production of a string via crossover, by looking at its constituent schemata, i.e. by using the Building Block basis. Suppose that we only have one-point crossover and no selection or mutation. We adopt the following notation, given that γ is a schema:

- $D(\gamma)$ is the set of indices for which γ has defined bit values. For example

$$D(1 ** 1 1 *) = \{1, 4, 5\}$$

- $L_j(\gamma)$ is the schema which has the same defining bits as γ for all indices $\leq j$, and stars elsewhere. For example, $L_4(1 ** 1 1 *) = 1 ** 1 **$.
- $R_j(\gamma)$ is the schema which has the same defining bits as γ for all indices $> j$, and stars elsewhere. For example, $R_4(1 ** 1 1 *) = ** ** 1 *$.
- If the population at time t is $p \in \Lambda$, we write $P(\gamma, t) = \sum_{i \in \gamma} p_i$.

Then, following [Stephens and Waelbroek, 1997] we have

$$P(\gamma, t + 1) = \sum_{j \in D(\gamma)} P(L_j(\gamma), t) P(R_j(\gamma), t)$$

Notice that this equation also applies to strings, by associating a string with the schema having all the corresponding bits defined.

We can use this equation to see how a given string can be created by crossover over several generations. For example, the string 1 1 1 can be created from the pairs 1 * *, * 1 1 and 1 1 *, * * 1. In the previous generation the schema * 1 1 may have been created from the pair * 1 *, * * 1. We see that there are several different possible “family trees” that can be constructed, the leaves of which are the order-one schemata which match the string at the root. All of the elements of the trees however are elements of the Building Block basis associated with the string of interest.

5 Mutation and the Genotype-Phenotype coarse-graining

The unitation coarse graining is a particular case of the genotype-phenotype coarse-graining in which the fitness of a string is its Hamming weight (or the Hamming distance to the 0 string, $f(i) = w(i) = d(i, 0)$). Now we can write the equivalence classes like $\gamma_j = \{i \in \Omega | w(i) = j\}$. In this particular case the reduction of the search space is huge, going from 2^ℓ degrees of freedom to only $\ell + 1$ effective degrees of freedom.

As an example of the coarse-graining technique in this scenario, consider a GA with probability of crossover zero, probability of mutation μ and $\ell = 2$, with the one-max fitness function. The set of equations describing the evolution of the system are as follows

$$\begin{pmatrix} p_{00}(t+1) \\ p_{01}(t+1) \\ p_{10}(t+1) \\ p_{11}(t+1) \end{pmatrix} = \begin{pmatrix} (1-\mu)^2 & (1-\mu)\mu & (1-\mu)\mu & \mu^2 \\ (1-\mu)\mu & (1-\mu)^2 & \mu^2 & (1-\mu)\mu \\ (1-\mu)\mu & \mu^2 & (1-\mu)^2 & (1-\mu)\mu \\ \mu^2 & (1-\mu)\mu & (1-\mu)\mu & (1-\mu)^2 \end{pmatrix} \begin{pmatrix} p_{00}(t) \\ p_{01}(t) \\ p_{10}(t) \\ p_{11}(t) \end{pmatrix} \quad (3)$$

$$\times \frac{1}{p_{01}(t) + p_{10}(t) + 2p_{11}(t)} \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 2 \end{pmatrix} \begin{pmatrix} p_{00}(t) \\ p_{01}(t) \\ p_{10}(t) \\ p_{11}(t) \end{pmatrix}$$

While in the coarse-grained basis the equations are:

$$\begin{pmatrix} p(\gamma_0, t+1) \\ p(\gamma_1, t+1) \\ p(\gamma_2, t+1) \end{pmatrix} = \begin{pmatrix} (1-\mu)^2 & (1-\mu)\mu & \mu^2 \\ 2(1-\mu)\mu & (1-\mu)^2 + \mu^2 & 2(1-\mu)\mu \\ \mu^2 & (1-\mu)\mu & (1-\mu)^2 \end{pmatrix} \begin{pmatrix} p(\gamma_0, t) \\ p(\gamma_1, t) \\ p(\gamma_2, t) \end{pmatrix} \quad (4)$$

$$\times \frac{1}{p(\gamma_1, t) + 2p(\gamma_2, t)} \begin{pmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 2 \end{pmatrix} \begin{pmatrix} p(\gamma_0, t) \\ p(\gamma_1, t) \\ p(\gamma_2, t) \end{pmatrix}$$

We can solve this system using a similarity transformation for the state-transition matrix, finding the eigenvectors and replacing the original matrix with a diagonal similar matrix (see, for example, chapter 6 of [Reeves and Rowe, 2001]). The results are shown in figure 1, where we can see the fixed points as a function of the probability of mutation irrespective of the initial conditions of the population.

The unitation coarse-graining allow us to eliminate a redundant variable. Of course, it proves to be more useful as we increase the dimension of the search space [Rowe, 1998].

In contrast, for the Eigen model the genotype-phenotype coarse-graining is not compatible with mutation. Instead, we divide the space into Hamming distance classes from the master sequence. That is $\gamma_j = \{i \in \Omega | d(i, c_{ms}) = j\}$, where c_{ms} is the master sequence. In the case where $c_{ms} = 0$, this gives us the unitation coarse-graining.

Now we can solve the system using the same method as before, considering $f(c_{ms}) \gg f(j), \forall j \neq c_{ms}$. The results in the long-time limit for the two-bit problem are shown in figure 2. Note that qualitatively we obtain a very similar behaviour to the onemax landscape. This means that in spite of the fact that all the strings different to the master sequence have the same low fitness, evolution favours those close to the master sequence

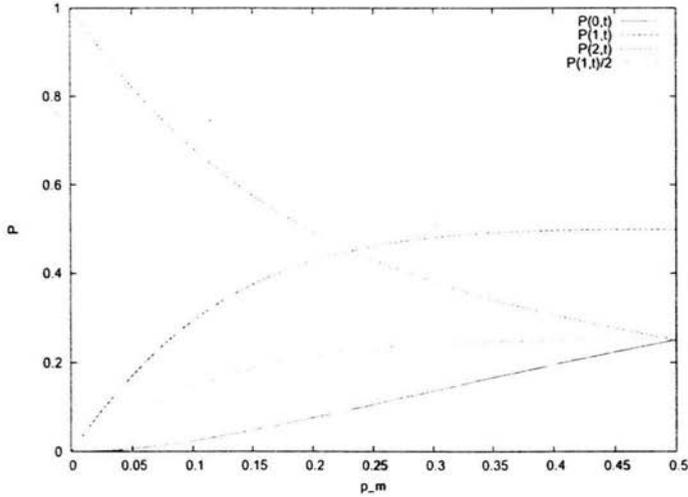


Fig. 1. Asymptotic string proportions in a onemax landscape with mutation rate p_m and no crossover. $P(2)$ decreases as we increase the probability of mutation, while $P(1)$ and $P(0)$ increase with p_m , and when p_m reaches $1/2$, all the strings are equally represented.

as we can see in figure 2. This phenomenon is known in the literature as the formation of a quasi-species.

6 Linkage-disequilibrium

Let ij be any string of 2 bits, $i, j \in \{0, 1\}$, then the dynamics of the system under proportional selection (with the onemax landscape) and one-point crossover is given by the equations:

$$\begin{aligned}
 P(ij, t+1) &= (1 - p_c)P'(ij, t) + p_c P'(ij, t) [P'(ij, t) + P'(i \bar{j}, t) + P'(\bar{i} j, t)] \\
 &\quad + p_c P'(i \bar{j}, t) P'(\bar{i} j, t) \\
 &= (1 - p_c)P'(ij, t) + p_c P'(ij, t) [1 - P'(\bar{i} \bar{j}, t)] + p_c P'(i \bar{j}, t) P'(\bar{i} j, t) \\
 &= P'(ij, t) + p_c [P'(i \bar{j}, t) P'(\bar{i} j, t) - P'(ij, t) P'(\bar{i} \bar{j}, t)]
 \end{aligned}$$

where P' is the proportion after selection, p_c is the probability of crossover and \bar{i} is the complement base 2 of i .

$\Delta'(t) = p_c [P'(i \bar{j}, t) P'(\bar{i} j, t) - P'(ij, t) P'(\bar{i} \bar{j}, t)]$ is the Selection Weighted Linkage Disequilibrium Coefficient (SWLDC) explicitly introduced in [Stephens, 2001] (and implicit in earlier work) in analogy with the original Linkage Disequilibrium Coefficient (LDC), well known in population biology, that measures how far is the current population from Robbins proportions. We can write

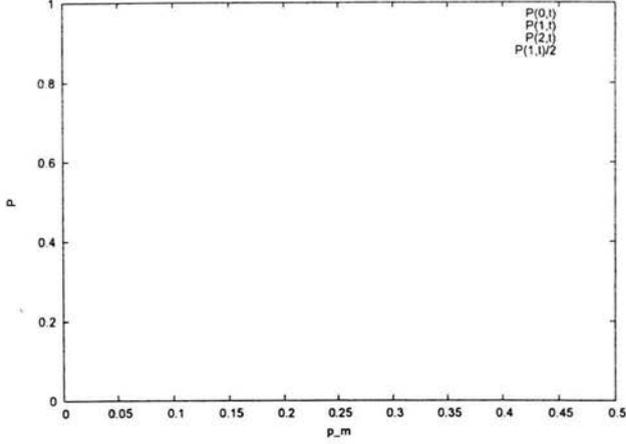


Fig. 2. Asymptotic string proportions in a Eigen model landscape with mutation rate p_m and no crossover.

$$w(t+1) = \frac{p_c x(t)y(t)}{\bar{f}^2(t)} \quad (5)$$

$$x(t+1) = \frac{x(t)}{\bar{f}(t)} - \frac{p_c x(t)y(t)}{\bar{f}^2(t)} \quad (6)$$

$$y(t+1) = \frac{y(t)}{\bar{f}(t)} - \frac{p_c x(t)y(t)}{\bar{f}^2(t)} \quad (7)$$

$$z(t+1) = \frac{2z(t)}{\bar{f}(t)} + \frac{p_c x(t)y(t)}{\bar{f}^2(t)} \quad (8)$$

where $w(t) = P(00, t)$, $x(t) = P(01, t)$, $y(t) = P(10, t)$, $z(t) = P(11, t)$, notice that $P(00, t) = \Delta'(t) \forall t > 0$

Substituting the value of $x(t)$ from (6) and $y(t)$ from (7) in (5)

$$w(t+1) = \frac{p_c}{\bar{f}^2(t)} \left[\frac{x(t-1)}{\bar{f}(t-1)} - \frac{p_c x(t-1)y(t-1)}{\bar{f}^2(t-1)} \right] \left[\frac{y(t-1)}{\bar{f}(t-1)} - \frac{p_c x(t-1)y(t-1)}{\bar{f}^2(t-1)} \right] \quad (9)$$

which we can rewrite as:

$$w(t+1) = \frac{1}{\bar{f}^2(t)} \left[w(t) - p_c w(t) \left(\frac{x(t-1) + y(t-1)}{\bar{f}(t-1)} \right) + p_c w^2(t) \right] \quad (10)$$

As there are four genotypes and three phenotypes we now coarse-grain using the genotype-phenotype map, considering the phenotypic variable $b(t) = x(t) + y(t)$, adding (6) and (7) at time t to find

$$\frac{b(t-1)}{\bar{f}(t-1)} = b(t) + 2w(t) \quad (11)$$

which can then be substituted into (10) to obtain

$$w(t+1) = \frac{w(t)}{f^2(t)} [1 - p_c(w(t) + b(t))] \quad (12)$$

with $\bar{f}(t) = b(t) + 2z(t)$. Now if we assume $p_c = 1$ we can simplify (12) to

$$w(t+1) = \frac{w(t)}{2z(t) + b(t)} \left[\frac{z(t)}{2z(t) + b(t)} \right] \quad (13)$$

Note that this equation, because of the substitution (11), is only valid for $w(t) = 0$, as the latter is a fixed point of (13) but not of (5) (unless $x(t) = y(t) = 0$ as well), and also for $p_c > 0$.

Let $g(t) = \left[\frac{z(t)}{2z(t) + b(t)} \right]$, iterating we find the following solution

$$w(t) = \frac{w(0) \prod_{i=0}^{t-1} g(i)}{2^t z(0) + b(0) + \delta_{t \geq 3} w(0) \sum_{i=1}^{t-2} 2^i \prod_{j=0}^{t-2-i} g(j)} \quad (14)$$

where $\delta_C = 1$ if C is true and 0 otherwise. It is easy to see that $g(t) \leq \frac{1}{2} \forall t$, so we can write

$$\begin{aligned} w(t) &\leq \frac{\left(\frac{1}{2}\right)^t w(0)}{2^t z(0) + b(0) + \delta_{t \geq 3} w(0) \left(\sum_{i=1}^{t-2} 2^i \left(\frac{1}{2}\right)^{t-i-1}\right)} \\ &= \frac{\left(\frac{1}{2}\right)^t w(0)}{2^t z(0) + b(0) + \delta_{t \geq 3} \frac{w(0)}{3} (2^{t-1} - 2^{3-t})} \\ &\leq \frac{\left(\frac{1}{2}\right)^t w(0)}{2^t z(0) + b(0)} = w^*(t) \end{aligned} \quad (15)$$

A comparison of the bound and the actual results are shown in fig. 3. Note that we would expect the bound to be better when $b(t) \ll z(t)$, however, for 3 $b(0) = 2z(0)$ so it is interesting to see that the bound gives reasonable results. Of course, strictly speaking the derived bound is for $p_c = 1$. The weaker the selective difference between 11 and 10 or 01 then the worse the bound. On the contrary, when the selective advantage of 11 over the other strings is large we expect the bound to become better and better. A further simplification lead us to

$$w^*(t) = \frac{\left(\frac{1}{2}\right)^t w(0)}{2^t z(0) + b(0)} \leq \left(\frac{1}{4}\right)^t \frac{w(0)}{z(0)} \quad (16)$$

where we can appreciate more clearly the exponential decay of the SWLDC.

By the definition of $w(t)$ in (5) and its value in (12) when $p_c = 1$ we can conclude that in every generation before selection we have $w(t)z(t) - x(t)y(t) = 0$ i.e. the population is in linkage (in the usual sense). This means that the population follows a path (in the simplex) always in linkage equilibrium, i.e. on the Geiringer manifold,

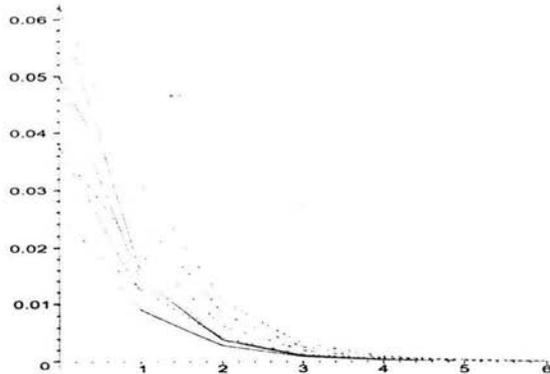


Fig. 3. Linkage Disequilibrium Coefficient $w(t)$ (continuous line) and the bound $w^*(t)$ (dotted line) in a typical run with a random initial population. The initial values correspond to different values of $p_c = \frac{j+2}{5}, j \in \{0, 1, 2, 3\}$

towards a selective linkage equilibrium. Clearly, the latter is of more relevance than the former for the dynamical evolution.

In figure (4) we can see the different “contours” with the same SWLDC as a function of the coarse-grained variable $b(t)$. Notice $b(t) = P(1, t)$ plays an important role in the solution of the equations suggesting again a possible effective degree of freedom.

As the effects of crossover decay in time, as $b(t)$ gets smaller, we might think about approximating the usual iteration sequence SCSCSC... with the sequence SCSSSS.... Where S denotes selection and C crossover. The effects can be seen in Figure 5. Note how the approximation to SCSCSC... uniformly improves in a “perturbative” fashion as we include more Cs in the sequence, four Cs giving a very good approximation.

7 Conclusion

We have discussed the notion of coarse graining in GAs, giving a formal definition and some representative examples such as genotype-phenotype mappings, schemata, error classes. We discussed how the dynamical equations for a GA transform under a coarse graining comparing and contrasting the notions of form invariance and compatibility and discussed some practical issues and potential problems that arise when applying arbitrary coarse-grainings. However, the evolution equations of certain GAs can be simplified with an appropriate choice of subsets. Schemata are natural subsets to consider as the dynamics is form invariant under a schema coarse graining for selection, mutation and crossover, whereas only mutation and crossover are compatible with the dynamics. Unitation classes were also seen to be natural in the case of selection and bitwise mutation. Genotype-phenotype coarse-grainings are natural in the case of selection, since they give rise to constant (that is, static) fitness values for the subsets. We have illustrated these ideas and problems with some simple examples. Finally, we

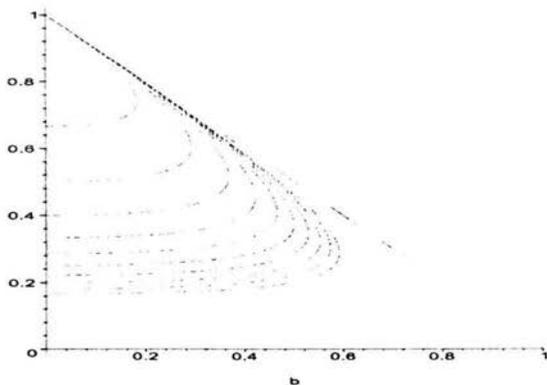


Fig. 4. The dotted lines represent zones with similar SLDC in the simplex as a function of $x(t) + y(t)$. The continuous line is the limit of the simplex.

have shown how a genotype-phenotype coarse-graining can help calculate an estimate for the selective linkage disequilibrium coefficient in the two-bit case.

8 Acknowledgements

This work was partially done while AAC was a visitor at the University of Birmingham, partially sponsored by a DGEP-UNAM grant. AAC would like to thank Jon Rowe and Xin Yao for all their attention and help to make this exchange possible. CRS acknowledges support from DGAPA project ES100201 and Conacyt project 30422-E.

References

- [Eigen, 1971] Eigen, M. (1971). Self-organization of matter and the evolution of biological macromolecules. *Naturwissenschaften*, 58:465.
- [Holland, 1975] Holland, J. (1975). *Adaptation in Natural and Artificial Systems*. University of Michigan Press, Ann Arbor, Michigan.
- [Radcliffe, 1992] Radcliffe, N. J. (1992). The algebra of genetic algorithms. *Anal. of Mathematics and Artificial Intelligence*, 10:339–384.
- [Reeves and Rowe, 2001] Reeves, C. R. and Rowe, J. E. (2001). *Genetic Algorithms — Principles and Perspectives*, chapter 6, The Dynamical Systems Model. Kluwer Academic Publishers.
- [Rowe, 1998] Rowe, J. (1998). Population fixed-points for functions of unitation. In Banzhaf, W. and Reeves, C., editors, *Foundations of Genetic Algorithms*, pages 69–84. Morgan Kaufmann.
- [Stadler and Stephens, 2003] Stadler, P. F. and Stephens, C. R. (2003). Landscapes and effective fitness. *Comm. Theor. Biol.* Accepted for publication.
- [Stephens, 2001] Stephens, C. R. (2001). Some exact results from a coarse grained formulation of genetic dynamics. In Spector, L., Goodman, E. D., Wu, A., Langdon, W. B., Voigt, H.-M., Gen, M., Sen, S., Dorigo, M., Pezeshk, S., Garzon, M. H., and Burke, E., editors, *Proceedings of the Genetic and Evolutionary Computation Conference (GECCO-2001)*, pages 631–638, San Francisco, California, USA. Morgan Kaufmann.

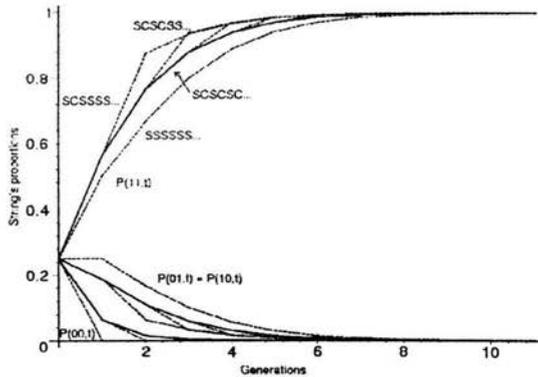


Fig. 5. The dotted lines represent the SCSSSS and the continuous SCSCSCSC

- [Stephens, 2003] Stephens, C. R. (2003). The renormalization group and the dynamics of genetic systems. *Acta Phys. Slov.*, 52:515–524.
- [Stephens and Waelbroek, 1997] Stephens, C. R. and Waelbroek, H. (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 Kaufman.
- [van Nimwegen et al., 1997] van Nimwegen, E., Crutchfield, J. P., and Mitchell, M. (1997). Finite populations induce metastability in evolutionary search. *Physics Letters A*, 229:144–150.
- [Vose, 1991] Vose, M. D. (1991). Generalizing the notion of a schema in genetic algorithms. *Artificial Intelligence*, 50:385–396.
- [Vose, 1999] Vose, M. D. (1999). *The Simple Genetic Algorithm: Foundations and Theory*. MIT Press, Cambridge, MA.
- [Vose and Wright, 2001] Vose, M. D. and Wright, A. H. (2001). Form invariance and implicit parallelism. *Evolutionary Computation*, 9(3):355–370.

Bibliografía

- [1] Fisher R. A. *The Genetical Theory of Natural Selection*. Oxford University Press, 1930.
- [2] L. Altenberg. Nk fitness landscapes. In *Handbook of Evolutionary Computation*, page B2.7.2. Oxford University Press, 1997.
- [3] Lee Altenberg. The schema theorem and price's theorem. In Darrell Whitley and Michael Vose, editors, *Foundations of Genetic Algorithms*, volume 3, pages 23–49. Morgan Kaufmann, San Francisco, 1995.
- [4] R. Burger. *The Mathematical Theory of Selection, Recombination and Mutation*. Wiley, 2000.
- [5] Andrés Aguilar Contreras, Jonattan E. Rowe, and Chris Stephens. Coarse-graining in genetic algorithms: some issues and examples. A ser publicado en GECCO, 2003.
- [6] M. Eigen. Self-organization of matter and the evolution of biological macromolecules. *Naturwissenschaften*, 58:465, 1971.
- [7] David B. Fogel. *Evolutionary Computation The Fossil Record*. IEEE Press, 1998.
- [8] H. Geiringer. On the probability theory of linkage in mendelian heredity. *Annals of Mathematical Statistics*, 15:25–27, 1944.
- [9] John H. Gillespie. *Population Genetics A Concise Guide*. The Johns Hopkins University Press, 1998.
- [10] D. E. Goldberg. *Genetic Algorithms in Search, Optimization and Machine Learning*. Addison-Wesley, 1989.
- [11] John Holland. *Adaptation in Natural and Artificial Systems*. University of Michigan Press, Ann Arbor, Michigan, 1975.
- [12] S. A. Kauffman. *The origins of Order*. Oxford University Press, 1993.
- [13] M. Kimura. *The neutral theory of molecular evolution*, pages 208–233. Sinauer Associates, Sunderland, Mass, 1983.

- [14] Nix and Vose M. Modeling genetic algorithms with markov chains. *Annals of Mathematics and Artificial Intelligence*, (5):79–88, 1992.
- [15] R. Poli. Hyperschema theory for GP with one point crossover, building blocks and some new results in GA theory. In Riccardo Poli, Wolfgang Banzhaf, William B. Langdon, Julian F. Miller, Peter Nordin, and Terence C. Fogarty, editors, *Genetic Programming Proceedings of EuroGP'2000*, volume 1802, pages 163–180, Edinburgh, 2000. Springer Verlag.
- [16] R. Poli, C. R. Stephens, A. Wright, and J. Rowe. A schema theory based extension of Geiringer's theorem for linear GP and variable length GAs under homologous crossover. In Jonathan Rowe, Riccardo Poli, and Kenneth A. De Jong, editors, *Foundations of Genetic Algorithms VII*, Torremolinos, Spain, 4-6 September 2002. Morgan Kaufmann. Accepted.
- [17] A. Prügel-Bennett and A. Rogers. Modelling finite populations. FOGA talk 2002.
- [18] A. Prügel-Bennett and A. Rogers. Modelling GA dynamics. In L. Kallel, B.Naudts, and A. Rogers, editors, *Proceedings Theoretical Aspects of Evolutionary Computing*, pages 59–86, 2001.
- [19] Colin R. Reeves. Fitness landscapes: A guided tour. PPSN2000 tutorial, School of Mathematical and Information Sciences Coventry University, 2000.
- [20] Colin R. Reeves and Jonathan E. Rowe. *Genetic Algorithms-Principles and Perspectives A guide to GA Theory*, chapter 6 The Dynamical Systems Model. Kluwer Academic Publishers, 2001.
- [21] J Rowe. Population fixed-points for functions of unitation. In Wolfgang Banzhaf and Colin Reeves, editors, *Foundations of Genetic Algorithms*, volume 5, pages 69–84. Morgan Kaufmann, 1998.
- [22] Peter F. Stadler and Christopher R. Stephens. Landscapes and effective fitness, 2002. A ser publicado en Comm. Theor. Biol.
- [23] C. R. Stephens. Some exact results from a coarse grained formulation of genetic dynamics. In Lee Spector, Erik D. Goodman, Annie Wu, W.B. Langdon, Hans-Michael Voigt, Mitsuo Gen, Sandip Sen, Marco Dorigo, Shahram Pezeshk, Max H. Garzon, and Edmund Burke, editors, *Proceedings of the Genetic and Evolutionary Computation Conference (GECCO-2001)*, pages 631–638, San Francisco, California, USA, 7-11 July 2001. Morgan Kaufmann.
- [24] C. R. Stephens and J. Mora Vargas. Effective fitness as an alternative paradigm for evolutionary computation I: General formalism. *Genetic Programming and Evolvable Machines*, pages 363–378, 2000.

- [25] C. R. Stephens and J. Mora Vargas. Effective fitness as an alternative paradigm for evolutionary computation II: Examples and applications. *Genetic Programming and Evolvable Machines*, pages 363–378, 2000.
- [26] C. R. Stephens and H. Waelbroek. 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 Kaufman, 1997.
- [27] C. R. Stephens and H. Waelbroek. Analysis of the effective degrees of freedom in genetic algorithms. *Physical Review*, 1998.
- [28] Chris Stephens and Henri Waelbroek. Schema evolution and building blocks. *Evolutionary Computation*, 7(2):109–124, 1999.
- [29] Monroe W. Strickberger. *Evolution*. Jones and Bartlett Publishers, 1990.
- [30] Michael. D. Vose. *The Simple Genetic Algorithm: Foundations and Theory*. The MIT Press, Cambridge, MA., 1999.
- [31] Darrell Whitley. A genetic algorithm tutorial. Technical Report CS-93-103, Colorado State University, March 1993.
- [32] L. Darrell Whitley. *Foundations of Genetic Algorithms*, volume 2. Morgan Kaufman Publishers, 1993.
- [33] Sewell Wright. The roles of mutation, inbreeding, cross breeding and selection in evolution. In D. F. Jones, editor, *Proceedings of the 6th International Congress on Genetics*, volume 1, pages 356–366, 1932.