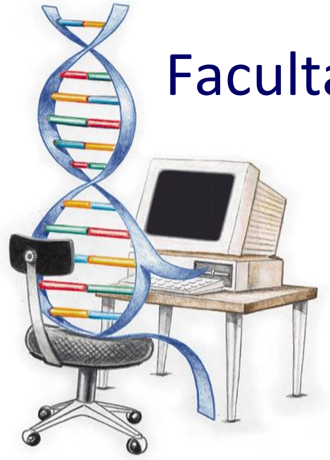
# ALGORITMOS EVOLUTIVOS
# Curso 2024

## Tema 7: Fundamentos teóricos



Centro de Cálculo, Instituto de Computación
Facultad de Ingeniería, Universidad de la República, Uruguay

# Fundamentos teóricos



- ¿Por qué funcionan los algoritmos evolutivos?
  - Teorema de los esquemas y procesamiento de esquemas
  - Hipótesis y falacia de los building blocks
- ¿Cómo trabaja un AE?
- Críticas y refinamientos al teorema de los esquemas
- Apuntes de Whitley sobre el funcionamiento de los AG
  - La importancia del operador de cruzamiento
  - La importancia del alfabeto utilizado
  - Esquemas procesados
  - El teorema de los esquemas revisado
- Análisis mediante cadenas de Markov

# AE: Fundamentos teóricos

## ¿ Por qué funcionan los AE ?

- Existen dos enfoques para analizar los fundamentos matemáticos de los algoritmos evolutivos:
  1. Teoría de los esquemas (*schema theorem*)
     - Es el enfoque seguido por Holland y Goldberg y se caracteriza por contabilizar el mecanismo de muestreo de soluciones en el espacio de búsqueda de un problema durante el ciclo evolutivo
  2. Considerar al algoritmo evolutivo como un proceso markoviano
     - Consiste en estudiar la evolución como un proceso estocástico que relaciona P(t) con P(t+1)

# AE: Fundamentos teóricos

## Teoría de los esquemas

UNIVERSIDAD DE LA REPÚBLICA URUGUAY

- La utilización de los esquemas fue propuesta por Holland a inicios de la década de 1970
- Un esquema es un formato o patrón que especifica una propiedad de un conjunto de soluciones, por lo cual caracteriza parcialmente a una solución en el espacio de genotipos
- La idea consiste en extender el alfabeto de símbolos utilizado para la codificación ($\Sigma$) con el metacáracter “*”, que concuerda con cualquier elemento de $\Sigma$
- Se pasa a trabajar sobre el alfabeto extendido $\Sigma \cup \{*\}$
- Cuando se trabaja sobre la codificación binaria el esquema se define sobre el alfabeto extendido $\{0,1,*\}$

# Teoría de los esquemas

## Qué son los esquemas ?

UNIVERSIDAD DE LA REPÚBLICA URUGUAY

- Por ejemplo, para un largo de codificación l = 7 considérense los esquemas E1, E2 y E3:
  - E1: `01**1**`
  - E2: `**100*0`
  - E3: `*******`
  - La tira `0110100` concuerda con E1 y E3.
  - La tira `0010001` concuerda solamente con E3.
  - E2 expresa al conjunto de tiras {`0010000`, `0010010`, `0110000`, `0110010`, `1010000`, `1010010`, `1110000`, `1110010`}.
- Un esquema representa un hiperplano en el espacio multidimensional que representa un conjunto de soluciones con propiedades comunes.

# Teoría de los esquemas

## Qué son los esquemas ?

UNIVERSIDAD DE LA REPÚBLICA URUGUAY

- Para alfabetos de cardinalidad k y un espacio de dimensión l, existen $(k + 1)^l$ esquemas
- Por ejemplo, para l=5 y k=2, se tienen 32 posibles individuos y 243 posibles esquemas
- ¿ Cuál es la utilidad de los esquemas ?
- Los individuos en forma aislada no son útiles para el análisis, ya que compiten con otros individuos. Es necesario considerar las clases de individuos (esquemas), que permiten describir en forma precisa las similitudes entre los individuos de una población y estudiar su influencia sobre el proceso de búsqueda

# Teoría de los esquemas

## Qué son los esquemas ?

UNIVERSIDAD DE LA REPÚBLICA URUGUAY

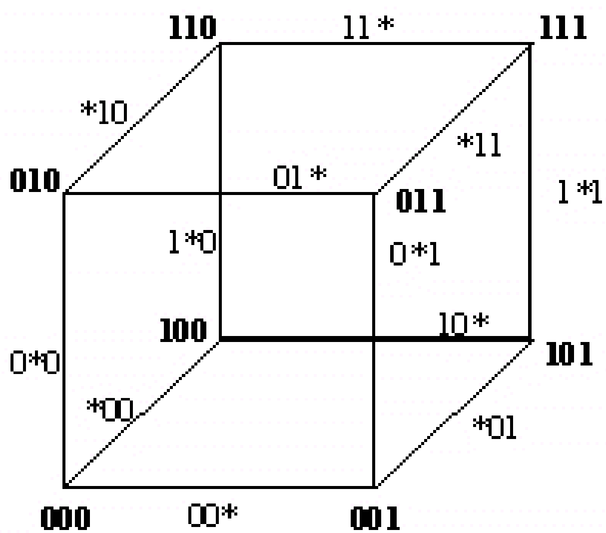
- En una codificación binaria cada individuo representa $2^l$ esquemas ya que cada posición puede contener el valor real o el metacarácter "*"
- Dependiendo de su diversidad, una población de tamaño n contendrá entre $2^l$ y $n*2^l$ esquemas
- Este rango es muy amplio, tan solo es una primera cota gruesa que se puede ajustar mediante los argumentos que se presentan a continuación

# Teoría de los esquemas

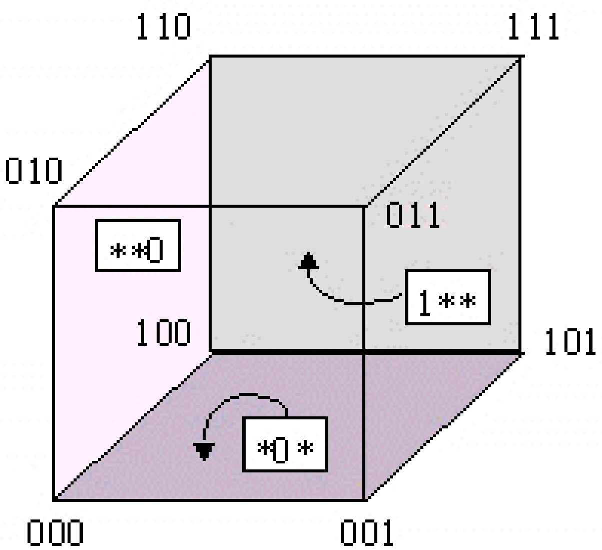
## Qué son los esquemas ?

- Para la presentación del teorema de los esquemas se requieren un par de definiciones previas:
  - *Orden de un esquema* H: o(H) es la cantidad de posiciones definidas en el esquema H
  - *Largo de definición de un esquema* H: $\delta$(H) es la distancia (en posiciones) entre el primer y el último valor definido en H, con lo cual $\delta$(H) pertenece al intervalo [0,l-1] siendo l el largo de la representación
- Por ejemplo:
  - E1: `011*1**` tiene o(E1) = 4 y $\delta$(E1) = 5 – 1 = 4
  - E2: `0******` tiene o(E2) = 1 y $\delta$(E2) = 0, ya que solamente tiene una posición definida

# Teoría de los esquemas




*Espacio de búsqueda para un problema con l=3 y k=2 (codificación binaria)*




*Tres de los seis esquemas de orden 1*

| orden del esquema | dimensión del hiperplano | interpretación geométrica | strings en el esquema | número de esquemas |
|---|---|---|---|---|
| 3 | 0 | punto | 1 | 8 |
| 2 | 1 | línea | 2 | 12 |
| 1 | 2 | plano | 4 | 6 |
| 0 | 3 | cubo | 8 | 1 |
| | | | total | **27** |

# AE: Fundamentos teóricos

## Teorema de los esquemas

- La idea detrás del análisis consiste en identificar similitudes entre los individuos de la población y relacionarlas con los valores altos de la función de fitness
- Para caracterizar estas similitudes se utilizan los esquemas
- El análisis se realiza sobre el algoritmo genético simple (Goldberg, 1989): se adopta como modelo teórico del funcionamiento del AG la selección proporcional, el operador de cruzamiento SPX y mutación de inversión de un bit
- Se considera en forma sucesiva el efecto de aplicar los operadores que trabajan en el ciclo evolutivo: la selección proporcional, el cruzamiento y la mutación

# AE: Fundamentos teóricos

## Teorema de los esquemas

- Sea m(H,t) el número de individuos representantes del esquema H en la generación t. Se estudia el efecto de aplicación de los operadores sobre m(H,t) sobre una población de n individuos
- Selección proporcional:
  - La selección proporcional copia individuos de acuerdo a los valores de fitness sobre el total de la población con probabilidad $p_i$

$$p_i = \frac{f(i)}{\sum_{j=1}^{j=n} f(j)}$$

# AE: Fundamentos teóricos

## Teorema de los esquemas

- El efecto de la selección proporcional para un individuo es

$$m(i,t+1) = m(i,t).n.\frac{f_i}{\sum_{j=1}^{n} f_j}$$

- Para un esquema resulta $m(H,t+1) = m(H,t).n.\frac{f(H)}{\sum_{j=1}^{n} f_j}$

  siendo $f(H)$ el valor promedio de fitness de individuos que representan al esquema H en el tiempo t

- El fitness promedio de la población es $\overline{f} = \frac{\sum_{j=1}^{n} f_j}{n}$

- El efecto de la selección proporcional sobre los esquemas se resume mediante la ecuación

$$\boxed{m(H,t+1) = m(H,t).\frac{f(H)}{\overline{f}}}$$

## Teorema de los esquemas

- El fitness del esquema *H* puede ser superior o inferior al fitness promedio
- Considerando $f(H) = \overline{f} + c\overline{f}$
  - Si c > 0 el fitness del esquema es superior al promedio
  - Si c < 0 el fitness del esquema es inferior al promedio
- Resulta $m(H,t+1) = m(H,t).\dfrac{\bar{f} + c.\bar{f}}{\bar{f}} = (1+c).m(H,t)$
- Propagando, se obtiene la siguiente ley exponencial:

$$\boxed{m(H,t+1) = (1+c)^t.m(H,0)}$$

# AE: Fundamentos teóricos

## Teorema de los esquemas

- El operador de selección proporcional opera asignando exponencialmente copias a aquellos individuos mejor adaptados y “extinguiendo” las características de individuos menos adaptados
- Los individuos mejor adaptados serán aquellos que concuerdan con esquemas cuyo fitness es superior al del promedio de la población

# AE: Fundamentos teóricos

## Teorema de los esquemas

- Analizando el efecto del cruzamiento, el operador resulta destructivo (*disruptivo*) para ciertos esquemas, por lo cual debe calcularse la probabilidad de supervivencia de un esquema
- Cuando se aplica SPX, un esquema sobrevive si el punto de corte cae fuera de su largo de definición
- Ejemplos
  - `10|**0`      el esquema se discontinúa
  - `10|***`      el esquema sobrevive

$\uparrow$

*punto de corte*

## Teorema de los esquemas

- Sea $P_S(H)$ la probabilidad de supervivencia del esquema H y $P_D(H)$ la probabilidad de disrupción del esquema H

$$P_S(H) = 1 - P_D(H)$$

- Resulta $P_S(H) \geq 1 - \dfrac{\delta(H)}{l-1}$

- La justificación para utilizar un "mayor o igual" y no un "igual" es porque existen casos donde un esquema sobrevive y no están siendo considerados. Por ejemplo, un esquema sobrevive si el individuo representante del esquema se cruza con otro que aporta la característica que el primero pierde

# AE: Fundamentos teóricos

## Teorema de los esquemas

- Sea $p_C$ la probabilidad de realizar el cruzamiento
- Incorporando a la ecuación precedente, la probabilidad de supervivencia de un esquema resulta:

$$P_S(H) \geq 1 - p_C . \frac{\delta(H)}{l-1}$$

- Combinando selección y cruzamiento, se obtiene:

$$\boxed{m(H, t+1) \geq m(H, t) \frac{f(H)}{\bar{f}} \left[ 1 - \left( p_C . \frac{\delta(H)}{l-1} \right) \right]}$$

- Se mantiene el crecimiento exponencial para aquellos esquemas con fitness superior al promedio y largo de definición bajo

# AE: Fundamentos teóricos

## Teorema de los esquemas

- Finalmente, debe incorporarse el efecto de la mutación
- Sea $p_M$ la probabilidad de mutación de un alelo
- Naturalmente la probabilidad de supervivencia de un alelo es $1-p_M$ con lo cual

$$P_S(H)=(1-p_M)^{O(H)}$$

- Aproximando por un desarrollo de Taylor y considerando valores de $p_M$ mucho más pequeños que 1 (como es habitual en los AE)

$$p_M << 1 \Rightarrow (1-p_M)^{O(H)} \approx 1-o(H)\,p_M$$

# AE: Fundamentos teóricos

## Teorema de los esquemas

$$\boxed{m(H,t+1) \geq m(H,t).\frac{f(H)}{\bar{f}}\left\{1-\left(p_C.\frac{\delta(H)}{l-1}\right)-o(H).p_M\right\}}$$

"Los esquemas cortos, de bajo largo de definición, cuyo fitness es superior al del promedio de la población, tienen una tasa de crecimiento exponencial. El AG hace este trabajo en un modo implícitamente paralelo para todos los esquemas de la población"

D. Goldberg, 1989

Genetic Algorithms in Search, Optimization, and Machine Learning

# AE: Fundamentos teóricos

## Procesamiento de esquemas

UNIVERSIDAD DE LA REPÚBLICA URUGUAY

- La idea es refinar la estimación de los esquemas que son procesados en una generación ( $n_S$ )
- Anteriormente se obtuvo la cota $2^l \leq n_S \leq n.2^l$ para una población de tamaño n de individuos de largo l
- Considerando los esquemas que sobreviven con probabilidad mayor a $p_S$ (constante), definidos por un largo $l_S < \varepsilon(l-1)+1$ siendo $\varepsilon < 1 - p_S$ una cierta tasa de error
- Dado ese largo del esquema es posible estimar una cota inferior de $n_S$ para aquellos esquemas útilmente procesados

**Autoestudio:**

**Procesamiento efectivo de esquemas**

Transparencias 22 a 25

Sección correspondiente en el texto de Golbderg.

# AE: Fundamentos teóricos

## Teorema de los esquemas

- Para realizarlo:
  - Se deberán contar los esquemas de largo $l_S$ o menor
  - Se deberá hallar un n "adecuado" para tener en promedio no más de un representante de esquemas de orden $l_S / 2$
- Veámoslo con un ejemplo:
  - l = 10: $\boxed{10111}$00010
  - Se contarán los esquemas en los primeros 5 lugares, de modo que la última posición quede fija
  - Se define el "metaesquema" # # # # 1*****, siendo # un "meta-metacáracter" que representa una posición fija (0 o 1) o un *
  - Como hay $l_S - 1$ posiciones para asignar, existirán $2^{(l_S - 1)}$ metaesquemas de este tipo

## Teorema de los esquemas



- Veámoslo con un ejemplo:
  - Para contar el total se “desplaza” la celda de 5 posiciones $(l - l_S + 1)$ veces
  - Por lo cual existirán $(l - l_S + 1).2^{(l_S - 1)}$ esquemas de largo $l_S$ o menos
- El conteo que se presentó es solamente para un individuo
- Generalizando para una población de tamaño n, una cota superior natural es $n.(l - l_S + 1).2^{(l_S - 1)}$ para esquemas de largo $l_S$ o menor.

# AE: Fundamentos teóricos

## Teorema de los esquemas

- Para refinar la estimación, considérese un tamaño de población $n = 2^{(l_S/2)}$
- Mediante esta elección del tamaño de la población se busca tener a lo sumo un representante de los esquemas de orden $l_S/2$ o mayor
- Como el número de esquemas corresponde a una distribución binomial, la mitad será de mayor orden y la mitad de menor orden que $l_S/2$
- Considerando la mitad de orden mayor, se tiene

$$n_S \geq n\left(l - l_S + 1\right)2^{l_S - 2}$$

- Como $n = 2^{(l_S/2)}$, resulta $n_S = \dfrac{\left(l - l_S + 1\right)n^3}{4} \cong \boxed{Cn^3}$

## Teorema de los esquemas



- El número de esquemas procesados en una población es $O(n^3)$.
- A pesar de la disrupción de esquemas largos y de alto largo de definición por la aplicación del cruzamiento y la mutación, los AE procesan una gran cantidad de esquemas, aunque manejan una cantidad relativamente baja de individuos
- Utilizar un tamaño de población $n = 2^{(l_S/2)}$ es una condición fuerte, habitualmente se utiliza un número más pequeño de individuos
- Paralelismo implícito:
  - Capacidad del AE de explorar en forma simultánea múltiples secciones del espacio de búsqueda
  - Bajo las hipótesis presentadas es de orden cúbico en n (puede ser menor para poblaciones más pequeñas)

# AE: Fundamentos teóricos

## Hipótesis de los building blocks

UNIVERSIDAD DE LA REPÚBLICA URUGUAY

- El teorema de los esquemas permite concluir que los esquemas cortos, de calidad por encima de la media y con un número bajo de posiciones fijas reciben un número exponencialmente creciente de individuos a lo largo de las generaciones
- Los esquemas cortos, con fitness superior al promedio y largo de definición se denominan building blocks
- Los building blocks son los bloques básicos que el algoritmo genético combina para construir soluciones casi-óptimas para el problema de optimización

# AE: Fundamentos teóricos

## ¿ Cómo trabaja un AE ?

UNIVERSIDAD DE LA REPÚBLICA URUGUAY

- El mecanismo del AE muestrea los esquemas definidos sobre el espacio de búsqueda asignando mayor “esfuerzo” en estudiar aquellas regiones (hiperplanos) de mejor valor promedio de fitness
- El mecanismo de selección garantiza que a través de las generaciones se asignen de manera exponencial individuos a los mejores esquemas
- El operador de cruzamiento explora nuevas combinaciones de los building blocks para lograr individuos más aptos
- El operador de mutación puede introducir nuevo material genotípico en la población, ya sea porque este no existiera en la población inicial o porque se hubiera perdido previamente

# AE: Fundamentos teóricos

## Críticas y refinamientos al teorema de los esquemas

- El enfoque presentado es el que introdujeron Holland y Goldberg para intentar explicar el éxito de los AG en la práctica
- Sin embargo, en su formulación original, el planteo no cierra del todo …
- Por ejemplo, cuando se utilizan *problemas deceptivos* la idea de los building blocks no parece ser aplicable …
- Conceptos teóricos avanzados:
  - Refinamientos de Whitley al teorema de los esquemas
  - Críticas (la falacia de los building blocks)
  - Problemas deceptivos
  - Análisis mediante cadenas de Markov

## Análisis de Whitley sobre el funcionamiento de los AE

- En su artículo "A genetic algorithm tutorial", D. Whitley plantea algunas observaciones al desarrollo teórico inicialmente propuesto por Holland y Goldberg
- Los aspectos fundamentales que se cubren son:
  - estudia la utilización de otros operadores de cruzamiento (UX)
  - las ventajas y/o desventajas del alfabeto considerado
  - ajuste al teorema de los esquemas a partir de la incorporación de los conceptos de ganacia y pérdida
  - ajuste al número de esquemas procesados útilmente

# AE: Fundamentos teóricos

## Análisis de Whitley sobre el funcionamiento de los AE

- La importancia del operador de cruzamiento
- La propuesta de Holland y Goldberg solamente consideraba el operador de cruzamiento de un punto (SPX) y dejaba como "ejercicio" el cruzamiento de dos puntos (2PX) ☺
- Whitley señala algunos aspectos a considerar del operador de cruzamiento uniforme (UX)
- Para que un esquema sobreviva, el hijo que herede el primer bit definido debe heredar todas las restantes posiciones definidas $(o(H) - 1)$
    - Considerando equiprobabilidad de selección entre los padres (no hay dominancia) la probabilidad de discontinuar un esquema es $1 - (1/2)^{(o(H) - 1)}$. No tiene sesgo con respecto a $\delta(H)$

# AE: Análisis de Whitley

## Importancia del operador de cruzamiento

UNIVERSIDAD DE LA REPÚBLICA URUGUAY

- Cruzamiento uniforme
  - Generalmente es más destructivo que SPX
  - El mecanismo del UX reduce al mínimo los efectos de la epistasis (fenómeno que implica la modificación de la acción de un gen por otro u otros genes)
  - Sea h la diferencia entre dos padres (para codificación binaria, será la distancia de Hamming):
    - SPX puede generar $2(h-1)$ individuos (sin considerar a los propios padres)
    - 2PX puede generar $h^2 - h$ individuos (sin considerar a los propios padres)
    - UX puede generar $2^h - 2$ individuos (sin considerar a los propios padres)

# AE: Análisis de Whitley

## Importancia del operador de cruzamiento

UNIVERSIDAD DE LA REPÚBLICA URUGUAY

- La Figura 1 ilustra el cruzamiento entre los padres `0000` y `1111` mediante los operadores SPX y UX.

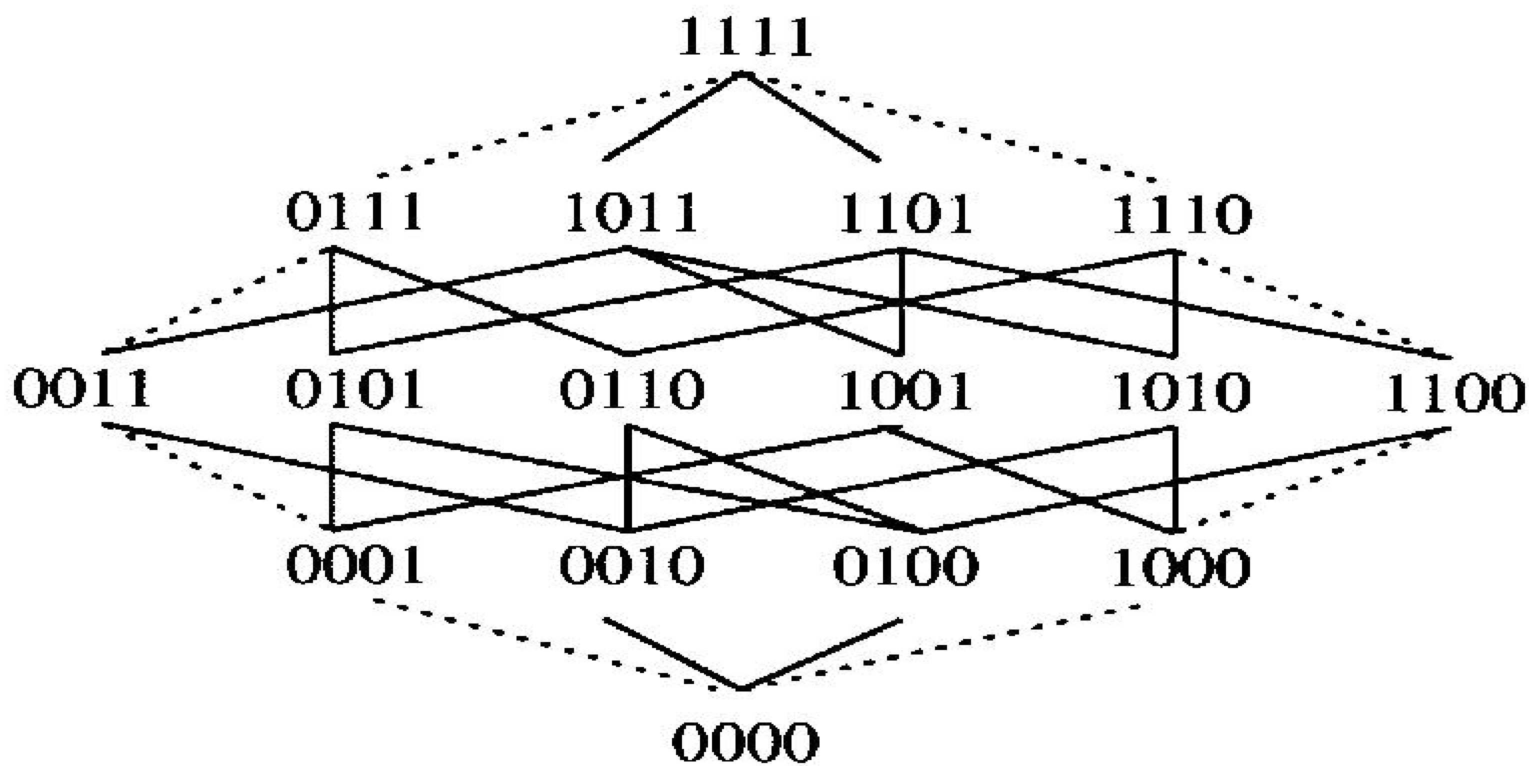



Figura 1: SPX solamente permite acceder a los punteados, UX a todos.

# AE: Análisis de Whitley

## Importancia del alfabeto utilizado

UNIVERSIDAD DE LA REPÚBLICA URUGUAY

- Alfabetos binarios
  - El alfabeto es minimal y maximiza el número de hiperplanos que son muestreados por la población
  - En contrapartida el largo de la representación usualmente es mayor
  - Cada esquema de orden 1 dividirá en dos el espacio de búsqueda
  - Hay 2*l* esquemas de largo 1
  - Para *o*(*H*) = 2 existen $C_2^l$ posibilidades para las posiciones fijas y $2^2$ posibles modos de asignar valores a ellas, resultando $2^2 C_2^l$ esquemas con *o*(*H*) = 2
  - En general existirán $2^i C_i^l$ esquemas de *o*(*H*) = *i*

# AE: Análisis de Whitley

## Importancia del alfabeto utilizado

- Alfabetos de cardinalidad mayor
  - Dado que existen más hiperplanos que en el caso binario, cada uno es muestreado con menor probabilidad que para cardinalidad 2 (utilizando una población con la misma cantidad de individuos).
  - Poder muestrear el espacio de búsqueda con la misma probabilidad que en el caso binario, el AE puede requerir más individuos y/o más generaciones.
  - Tiene como ventaja que permite el diseño de operadores específicos que incorporen algún tipo de semántica, por ejemplo para el manejo de permutaciones o estructuras complejas.
  - Además, el largo de la codificación será menor.

# AE: Análisis de Whitley

## Esquemas procesados

- Sea $\theta$ el orden más alto de hiperplanos que será representado en la población de tamaño *n* por al menos $\phi$ copias. Se requieren $\phi$ muestras de un hiperplano antes de asegurar que se está *muestreando en forma estadística* el hiperplano

$$\theta = \log\left(\frac{n}{\phi}\right)$$

- El número de esquemas de *o(H)* = $\theta$ es $2^{\theta} C^{l}_{\theta}$
- Es necesario demostrar que:

$$\left.\begin{array}{c} 2^{\theta} C^{l}_{\theta} \geq n^{3} \\ \theta = \log\left(\frac{n}{\phi}\right) \Rightarrow n = 2^{\theta}\phi \end{array}\right\} \Rightarrow 2^{\theta} C^{l}_{\theta} \geq \left(2^{\theta}\phi\right)^{3}$$

# AE: Análisis de Whitley

## Esquemas procesados

- Si se considera $l \geq 64$ y $2^6 \leq n \leq 2^{20}$, tomando $\phi = 8$ se muestrean "estadísticamente" esquemas de orden entre 3 y 17.
- Para ese caso particular de *l*, *n* y $\Phi$, la propiedad se cumple.
- Sin embargo, no se cumple genéricamente para cualquier valor de los parámetros.
- Si bien es más claro conceptualmente que el análisis de Goldberg, desde el punto de vista práctico sigue siendo poco útil (*n* sigue siendo muy grande)

# AE: Análisis de Whitley

## Teorema de los esquemas

UNIVERSIDAD DE LA REPÚBLICA URUGUAY

- Críticas de Whitley al enfoque de Goldberg
  - Es una visión cualitativamente adecuada, pero desde el punto de vista cuantitativo es poco útil
  - Se simplifica de sobremanera el efecto de los operadores al no considerarse el efecto de las ganancias y sobreestimarse las pérdidas, con lo cual se obtiene una cota inferior
  - La suposición que se realiza sobre c (el valor de “superioridad” sobre el promedio) como constante no se cumple en el largo plazo, ya que el fitness observado de un hiperplano H en un instante t varía a medida que las poblaciones se van moviendo en el espacio de búsqueda
  - Por estos motivos puede ser útil para estimar la cantidad de esquemas presentes en una o dos generaciones consecutivas pero no más allá

## Teorema de los esquemas

UNIVERSIDAD DE LA REPÚBLICA URUGUAY

- Se incorpora explícitamente el parámetro *t* a la función de fitness de los representantes de los esquemas
- Se incorpora el concepto de pérdidas y ganancias:

$$\boxed{m(H,t+1) = (1-p_c)m(H,t)\frac{f(H,t)}{\bar{f}} + p_c\left[m(H,t)\frac{f(H,t)}{\bar{f}}(1-losses)+gains\right]}$$

- Se plantea un modelo exacto para calcular el comportamiento de cada individuo (*Z*) del espacio de búsqueda

$$\boxed{P(Z,t+1) = P(Z,t)\frac{f(Z,t)}{\bar{f}}(1-\{p_c\, losses\}) + \{p_c\, gains\}}$$

# AE: Análisis de Whitley

## Teorema de los esquemas

UNIVERSIDAD DE LA REPÚBLICA URUGUAY

- Mediante el análisis de Whitley:
  - Se logra independizar el análisis del fitness observado en la generación, ya que el fitness de los strings es constante
  - Se pueden calcular las pérdidas y las ganancias en forma exacta.
  - Las pérdidas se generan cuando un individuo se cruza con otro y los hijos resultantes no preservan al individuo inicial
  - Las ganancias ocurren cuando dos individuos diferentes son cruzados creando una nueva copia de algún individuo
  - Se proponen ecuaciones para generar las pérdidas y las ganancias, pero son muy complejas y poco útiles en la práctica

# AE: Análisis de Whitley

## Teorema de los esquemas

- Ejemplo con ganancias y pérdidas
  - Sea el individuo $Z$ = `000`
  - $P_{I0}$ es la probabilidad de realizar el cruzamiento en cualquier punto de la tira (su valor es 1, pero se incorpora en la fórmula por claridad)
  - $P_{I1}$ es la probabilidad de realizar el cruzamiento entre el primer bit y el segundo
  - $P_{I2}$ es la probabilidad de realizar el cruzamiento entre el segundo bit y el tercero

## Teorema de los esquemas

- Ejemplo con ganancias y pérdidas para SPX sobre el individuo 000

$$losses = P_{I0}\frac{f(111)}{\bar{f}}P(111,t) + P_{I0}\frac{f(101)}{\bar{f}}P(101,t)$$
$$+ P_{I1}\frac{f(110)}{\bar{f}}P(110,t) + P_{I2}\frac{f(011)}{\bar{f}}P(011,t)$$

$$gains = P_{I0}\frac{f(001)}{\bar{f}}P(001,t)\frac{f(100)}{\bar{f}}P(100,t) + P_{I1}\frac{f(010)}{\bar{f}}P(010,t)\frac{f(100)}{\bar{f}}P(100,t)$$
$$+ P_{I1}\frac{f(011)}{\bar{f}}P(011,t)\frac{f(100)}{\bar{f}}P(100,t) + P_{I2}\frac{f(001)}{\bar{f}}P(001,t)\frac{f(110)}{\bar{f}}P(110,t)$$
$$+ P_{I2}\frac{f(001)}{\bar{f}}P(001,t)\frac{f(010)}{\bar{f}}P(010,t)$$

# AE: Fundamentos teóricos

## La falacia de los building blocks

UNIVERSIDAD DE LA REPÚBLICA URUGUAY

- El trabajo de Chris Thornton señaló las contradicciones existentes entre el teorema de los esquemas y la hipótesis de los building blocks
- El teorema de los esquemas establece que "*los esquemas cortos de bajo orden de definición y por encima del promedio tendrán una tasa de crecimiento exponencial*"
- La principal objeción planteada se vincula al hecho de que se asume que el fitness promedio para el esquema permanece constante durante el cruzamiento, es decir que el valor de fitness no es afectado por su contexto genético
- El teorema de los esquemas asume que la epístasis es baja
- En la práctica, esta condición depende del problema (en ciertos casos efectivamente es así y en otros no)

# AE: Fundamentos teóricos

## La falacia de los building blocks

- Considérese una población con valores de fitness entre 80 y 120, con promedio 100. La "ventaja" máxima posible de fitness es 20/100
- Cualquier esquema para el que el cociente entre el largo de definición y el largo exceda 20/100, no será conservado
- Si se consideran individuos de largo 100, el largo de definición deberá ser menor que 20
- La proporción de esquemas de 20 bits en relación a los posibles individuos es:

$$\frac{2^{20}}{2^{100}} * 100 = 0.827181 * 10^{-22}$$

- Se procesa una cantidad "despreciable" de esquemas
- Se viola el principio de que los esquemas deben ser cortos

# AE: Fundamentos teóricos

## La falacia de los building blocks

- La hipótesis de los building blocks establece que los esquemas cortos de bajo orden de definición (los building blocks) y buen fitness se combinan para formar esquemas con mejor fitness y mayor largo de definición
- *"The ability to produce fitter and fitter partial solutions by combining blocks is believed to be the primary source of the GA's search power"* (Forrest & Mitchell, 1996)
- La implicancia directa es que la epístasis genera un efecto positivo en el fitness
- Sin embargo, la combinación de building blocks generará building blocks más grandes con el paso de las generaciones. El tamaño de los esquemas ira aumentando

## La falacia de los building blocks



| Teorema de los esquemas | Hipótesis de los building blocks |
|---|---|
| Esquemas cortos | Tamaño creciente de los esquemas |
| Epístasis baja | Epístasis alta |

Tabla 1: Contradicciones en el enfoque teórico estándar de los AE (Thorton, 1997).

# AE: Fundamentos teóricos

## Otras consideraciones

- Se han presentado varias observaciones al análisis planteado por Goldberg, pero todavía hay más …
- Se asumen condiciones muy restrictivas: codificación binaria y cruce de un punto. Adicionalmente, en la forma en la que está formulado solamente es aplicable a problemas de optimización sin restricciones
- No garantiza la obtención del óptimo, solamente brinda una medida del número esperado de representantes de un esquema para el proceso evolutivo
- Puede suceder que un esquema que supuestamente tenga alta calidad en realidad sea de baja calidad ya que el muestreo de individuos se realiza con respecto a la media de la población

## Problemas deceptivos

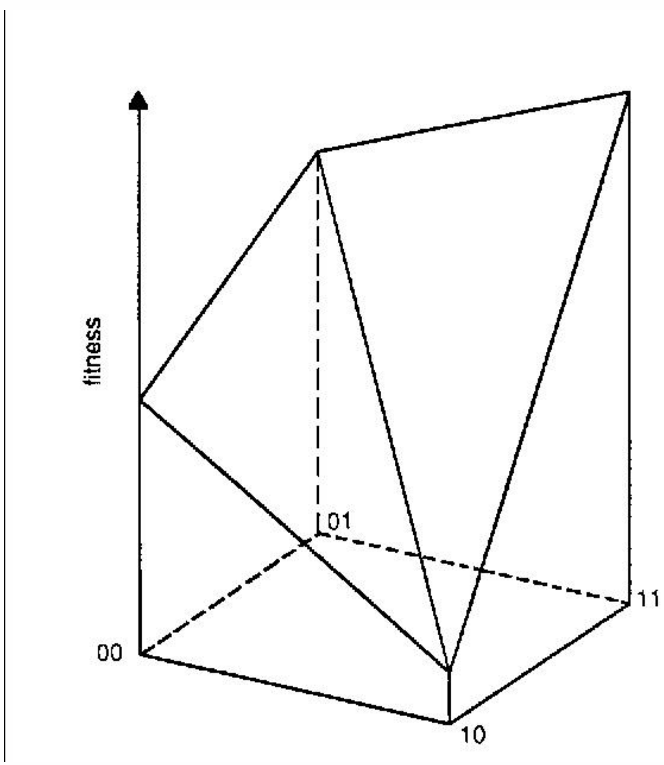
- Los AE tienen inconvenientes al resolver *problemas deceptivos*, (aquellos problemas en los cuales la combinación de esquemas cortos de gran calidad y de orden de definición bajo conduce a soluciones alejadas al óptimo)
- Por ejemplo, para el caso de individuos de 2 bits y siendo *f(11)* óptimo global:

$$\left.\begin{array}{l} f(0*) > f(1*) \\ f(*0) > f(*1) \end{array}\right\} \Rightarrow \begin{cases} \dfrac{f(00)+f(01)}{2} > \dfrac{f(10)+f(11)}{2} \\ \dfrac{f(00)+f(10)}{2} > \dfrac{f(01)+f(11)}{2} \end{cases}$$

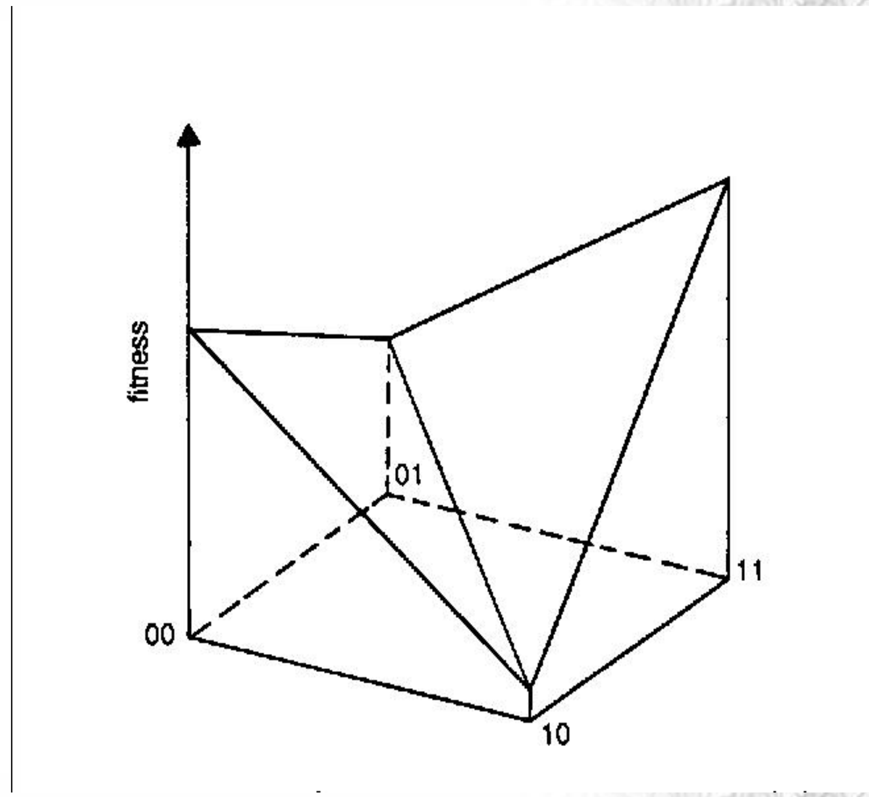
- Es imposible imponer ambas condiciones y que se mantenga *f(11)* como óptimo global
- Por ejemplo, si se impone *f(0*) > f(1*)*

# AE: Fundamentos teóricos

## Problemas deceptivos




(a) caso *f(01) > f(00)*




(b) caso *f(00)≥ f(01)*

Figura 2: Problemas deceptivos de dos bits con *máximo f(11)*

## Problemas deceptivos

- Un análisis completo de los problemas deceptivos puede verse en el artículo “Fundamental Principles of Deception in Genetic Search” (Whitley, 1991)
- Es evidente que la fundamentación presentada es útil desde el punto de vista cuantitativo, pero endeble formalmente
- Se necesita una fundamentación matemática más sólida que permita demostrar el correcto funcionamiento de AE como técnica de optimización
- El enfoque alternativo que se ha seguido para realizar el análisis de los algoritmos evolutivos ha sido la utilización de cadenas de Markov

# AE: Fundamentos teóricos

## Análisis mediante cadenas de Markov



- Algoritmo evolutivo considerado
  - Opera sobre una población de individuos con una función de fitness asociada
  - Se utiliza alguna regla de selección basada en la función de fitness para determinar los individuos que serán los padres de la nueva generación
  - Los nuevos individuos obtenidos a partir de los operadores de selección y cruzamiento sustituyen a los individuos de la población actual
  - No se restringe la codificación ni el operador de cruzamiento utilizado
  - El marco conceptual considerado es extensible a problemas con restricciones

## Análisis mediante cadenas de Markov



- Cadenas de Markov
  - Son procesos estocásticos invariantes en el tiempo y sin memoria
  - Para cada posible estado del proceso las probabilidades de transiciones entre los estados solamente dependen del estado actual
  - Se caracterizan mediante un vector $\pi$ y una matriz de transición P
  - El vector $\pi$ representa la probabilidad de que la cadena se encuentre inicialmente en un estado determinado
  - Cada elemento (i,j) de P representa la probabilidad de pasar del estado i al estado j en una única transición

- Cadenas de Markov
  - $\pi$P es un vector fila en el que cada componente es la probabilidad de estar en cada estado después de una transición.
  - $\pi P^k$ es un vector fila en el que cada componente es la probabilidad de estar en cada estado después de $k$ transiciones.
  - Dos estados *i* y *j* se comunican si al menos hay un camino de *i* a *j* y viceversa.
  - Se dice que un estado es absorbente si no se comunica con ningún otro estado.
  - Se dice que una cadena de Markov es irreductible si todos los estados se comunican entre sí.

# AE: Fundamentos teóricos

## Análisis mediante cadenas de Markov

- Para el análisis de los AE mediante cadenas de Markov se considera cada una de las posibles configuraciones de una población como estado de la cadena
- Por ejemplo, si se considera una codificación binaria y siendo *n* la cantidad de individuos de la población y *l* es el largo de cada individuo, el número de estados es $2^{nl}$
- Es claro que para casos realistas la cantidad de estados posibles es muy grande.
- Sin embargo, a pesar de tener una dimensión muy alta, el espacio de estados es finito, por lo cual se cumple que la cadena de Markov es finita

## Análisis mediante cadenas de Markov

UNIVERSIDAD DE LA REPÚBLICA URUGUAY

- *Algoritmo genético no elitista*
  - Un AG que solamente tenga selección y cruzamiento *no es irreductible* ya que tiene estados absorbentes
  - Los estados absorbentes serán aquellos en los que todos los miembros de la población son exactamente iguales
  - El resto de los estados serán estados transitorios
  - Se ha demostrado que para un número suficientemente grande de generaciones, la cadena llega con probabilidad 1 a un estado absorbente
  - Existe una probabilidad no nula de que el estado absorbente al que se converja no sea óptimo (no contenga al individuo óptimo)

## Análisis mediante cadenas de Markov



- Caso a): algoritmo evolutivo no elitista
  - Si se utiliza un operador de mutación, los estados considerados como absorbentes se transforman en transitorios
  - Si la probabilidad de mutación es muy baja predomina el efecto del operador de cruzamiento sobre el de mutación
  - Bajo estas hipótesis el conjunto de estados generados será muy parecido al conjunto de estados originalmente planteados
  - El operador de mutación solamente servirá para evitar la pérdida de información que puede impedir que se alcance la solución óptima

## Análisis mediante cadenas de Markov



- Caso a): algoritmo evolutivo elitista
  - Si se incorpora elitismo conservando la mejor solución encontrada a través de las generaciones, es posible demostrar que los AE convergen con probabilidad 1 a la solución óptima
  - Solamente hay un estado absorbente que incluye a todas las poblaciones que contienen a la solución óptima, mientras que el resto de los estados son transitorios
  - Se demuestra que para un número de generaciones infinitas siempre se llega al estado absorbente (convergencia asintótica), o equivalentemente se alcanza el óptimo global

# AE: Fundamentos teóricos

## Resumen: principales resultados teóricos

- Teorema de los esquemas
  - Formulación del procesamiento de soluciones mediante argumentos de conteo
  - Procesamiento de orden cúbico de esquemas en cada generación
  - Refinamiento de Whitley considerando ganancias y pérdidas
- Resultados contradictorios
  - La falacia de los building blocks
  - Problemas “difíciles” para los AE: problemas deceptivos
- Análisis mediante cadenas de Markov
  - Prueba teórica de convergencia global para un algoritmo evolutivo elitista