

Tiempo Esperado de Convergencia de un Algoritmo Genético con Parámetros Mínimos

María Margarita Reyes Sierra y Carlos A. Coello Coello
CINVESTAV-IPN

Grupo de Computación Evolutiva
Departamento de Ingeniería Eléctrica
Sección de Computación
Av. Instituto Politécnico Nacional No. 2508
Col. San Pedro Zacatenco
México D.F. 07300, MÉXICO
mreyes@computacion.cs.cinvestav.mx
ccoello@cs.cinvestav.mx

Resumen

El problema de caracterizar el comportamiento de los algoritmos genéticos en varios dominios es complejo puesto que varía con la aplicación así como con los parámetros con que se implemente. Se mostrará un trabajo basado en cadenas de Markov llevado a cabo para estimar el tiempo de convergencia de un algoritmo genético con parámetros mínimos, es decir, con una población de tamaño y longitud de sus individuos mínima.

1. Introducción

El Algoritmo Genético (AG) es una técnica *heurística* basada en la evolución natural, cuya aplicación en el campo de la optimización surge de la presencia de la selección natural en dicha teoría, que impone la supervivencia del más apto. En sus inicios, el AG (llamado ahora “clásico”) fue aplicado a problemas de optimización de funciones con un solo objetivo. En 1994 [7], fue publicada la demostración de la convergencia del Algoritmo Genético Elitista (AGE) al óptimo global de la función en cuestión, bajo ciertas condiciones.

En este trabajo se extiende el modelo correspondiente con el fin de estimar el tiempo esperado de convergencia de un AG con parámetros mínimos. Se detallan las características generales de la matriz de transición del AGE mencionadas en [7]. Los resultados obtenidos son comparados con los que presentan los experimentos correspondientes.

2. Antecedentes

En [1], Carol A. Ankenbrandt obtiene una cota para el tiempo de ejecución necesario para la convergencia del AG mediante un modelo basado en convergencia de alelos y una sencilla prueba de inducción matemática.

Por otra parte, dado que la selección natural usa la diversidad en una población para dar lugar a la adaptación, en [5] Louis & Rawlins desarrollan un modelo basado en distancias de Hamming y argumentan que, ignorando los efectos de la mutación, si no existe diversidad no hay nada en lo que la selección natural pueda trabajar, por lo que usan una medida de diversidad para estimar el tiempo de convergencia necesario.

Un estudio similar al que se presenta en [5] fue desarrollado por Goldberg y Segrest en 1987 [2], en el que el tiempo de convergencia de un AG cuya población consta de individuos de longitud 1 es estimado mediante un modelo markoviano. En el trabajo aquí presentado la longitud de los individuos considerados tiene longitud n .

3. Cadenas de Markov Finitas

3.1. Definiciones Básicas

El material presentado en esta sección fue tomado de las referencias [8, 3].

Definición 3.1 Si $\mathcal{S} \neq \emptyset$ es un conjunto finito y $\{X_t : t \in \mathbb{N}\}$ es una sucesión de variables aleatorias con valores en \mathcal{S} con la propiedad de que:

$$\begin{aligned} \mathbb{P}\{X_{t+1} = j | X_t = i, X_{t-1} = i_{t-1}, \dots, X_0 = i_0\} = \\ \mathbb{P}\{X_{t+1} = j | X_t = i\} =: p_{ij} \end{aligned}$$

para todo $t \geq 0$ y para todo $i, j \in \mathcal{S}$, entonces la sucesión $\{X_t : t \in \mathbb{N}\}$ es llamada una **cadena de Markov finita** con espacio de estados \mathcal{S} .

El número p_{ij} se llama **probabilidad de transición** del estado i al estado j en un paso. Como estamos suponiendo que dichas probabilidades son independientes de $t \in \mathbb{N}$, decimos que la cadena es **homogénea**.

Puesto que \mathcal{S} es finito, las probabilidades de transición pueden ser puestas en una **matriz de transición** $P = (p_{ij})_{i,j \in \mathcal{S}}$. Nótese que $\sum_j p_{ij} = 1$ para todo $i \in \mathcal{S}$.

El vector fila $\delta(t)$ con componentes $\delta_i(t) = \mathbb{P}\{X_t = i\}$ para todo $i \in \mathcal{S}$, denota la distribución de la cadena de Markov en el paso $t \geq 0$. Esta distribución puede calcularse iterativamente, pues

$$\delta(t) = \delta(t-1)P = \delta(0)P^t, \text{ para todo } t \geq 0.$$

Definición 3.2 Una matriz $P : n \times m$ es llamada **no negativa** ($P \geq 0$) si $p_{ij} \geq 0$ y **positiva** ($P > 0$) si $p_{ij} > 0$ para todo $i = 1, \dots, n$ y $j = 1, \dots, m$. Una matriz cuadrada no negativa es llamada **estocástica** si la suma de cada una de sus filas es igual a 1. Así, las matrices de transición son estocásticas.

Una matriz estocástica P es **primitiva** si

$$\exists k \in \mathbb{N} : P^k \text{ es positiva } (P^k > 0)$$

Por lo tanto, toda matriz positiva P es primitiva.

Una matriz P es **reducible** (por supuesto si no es irreducible y) si puede ser acomodada en la forma:

$$\begin{pmatrix} \mathbf{C} & \mathbf{0} \\ \mathbf{R} & \mathbf{T} \end{pmatrix}$$

con matrices cuadradas \mathbf{C} y \mathbf{T} .

Finalmente, una matriz estocástica P es **columna-permisible** si al menos un elemento en cada columna es positivo.

A continuación se presentan algunos resultados que serán de utilidad en secciones posteriores.

Lema 3.1 [3] Sean P , Q y R matrices estocásticas, donde Q es positiva y R es columna-permisible. Entonces la matriz producto PQR es positiva.

Teorema 3.1 [3] Sea $P_{n \times n}$ una matriz estocástica reducible, donde $C : m \times m$ es una matriz estocástica primitiva y $R, T \neq 0$. Entonces:

$$P^\infty = \lim_{k \rightarrow \infty} P^k = \lim_{k \rightarrow \infty} \begin{pmatrix} C^k & 0 \\ \sum_{i=0}^{k-1} T^i R C^{k-i} & T^k \end{pmatrix} = \begin{pmatrix} C^\infty & 0 \\ R_\infty & 0 \end{pmatrix}$$

es una matriz estocástica estable con $P^\infty = \mathbf{1}' p^\infty$, donde $\mathbf{1}'$ es un vector columna de unos y $p^\infty = p^0 P^\infty$ es única independientemente de la distribución inicial y, además, p^∞ satisface: $p_i^\infty > 0$ para $1 \leq i \leq m$ y $p_i^\infty = 0$ para $m < i \leq n$.

3.2. Cadenas Absorbentes

Considérese una cadena de Markov con espacio de estados finito $\mathcal{S} \neq \emptyset$. Los estados de una cadena de Markov se clasifican de acuerdo a si es posible ir de un determinado estado a otro estado determinado [4].

Definición 3.3 Decimos que el estado i **induce** al estado j y lo denotamos por $i \rightarrow j$ si y sólo si $p_{ij}^k > 0$ para algún $k \geq 1$. Si tenemos que $i \rightarrow j$ y $j \rightarrow i$ decimos que el estado i está **comunicado** con el estado j (o que los estados i, j son **comunicantes**) y lo denotamos por $i \leftrightarrow j$.

A partir de la definición anterior, los estados son divididos en “clases de equivalencia”. Dos estados están en la misma clase de equivalencia si están “comunicados”, es decir, si el proceso puede ir de un estado a otro y viceversa. Las clases de equivalencia son clasificadas como conjuntos **ergódicos** (también llamados **recurrentes**) o conjuntos **transitorios**. De esta manera los estados pertenecientes a ellas serán llamados estados ergódicos y transitorios, respectivamente.

Para cada cadena de Markov finita debe existir siempre al menos un conjunto ergódico; sin embargo, los conjuntos transitorios no necesariamente existen. Una vez que una cadena abandona un conjunto transitorio nunca regresará a él, mientras que una vez que entra en un conjunto ergódico nunca podrá abandonarlo.

En particular, si un conjunto ergódico contiene sólo un estado, éste recibe el nombre de estado **absorbente** pues una vez estando en él, la cadena de Markov se quedará allí por siempre. De lo anterior tenemos que un estado i es absorbente si y sólo si $p_{ii} = 1$ [4]. Ahora bien, una cadena con conjuntos transitorios tiene un comportamiento tal que se mueve hacia los conjuntos ergódicos. La probabilidad de que el proceso se encuentre dentro de un conjunto ergódico tiende a 1. Cuando todos los conjuntos ergódicos son conjuntos unitarios, este tipo de cadenas son llamadas **cadenas absorbentes**, pues eventualmente quedarán atrapadas en un estado absorbente.

Teorema 3.2 [4] *Para cualquier cadena de Markov finita absorbente, no importa el estado en el que el proceso inicie, la probabilidad de que el proceso se encuentre en un estado absorbente después de n pasos tiende a 1 conforme n tiende a infinito.*

Es importante considerar la forma canónica de la matriz de transición de una cadena de Markov [4]. Supongamos que tenemos s estados transitorios y $r - s$ estados ergódicos y que agrupamos a todos los conjuntos transitorios y a todos los conjuntos ergódicos, la forma es la siguiente:

$$P = \begin{pmatrix} & r-s & s \\ S & O \\ R & Q \end{pmatrix} \begin{matrix} \\ \} r-s \\ \} s \end{matrix}$$

La región O consiste completamente de ceros. La matriz $Q_{s \times s}$ representa a la cadena mientras ésta se encuentre en estados transitorios, la matriz $R_{s \times (r-s)}$ representa la transición de estados transitorios a estados ergódicos y la matriz $S_{(r-s) \times (r-s)}$ representa la cadena una vez que ha alcanzado un estado ergódico. Si consideramos una cadena absorbente, tenemos que por definición $S = I_{(r-s) \times (r-s)}$, así su forma canónica es:

$$P = \begin{pmatrix} & r-s & s \\ I & O \\ R & Q \end{pmatrix} \begin{matrix} \\ \} r-s \\ \} s \end{matrix}$$

Por el Teorema 3.1 podemos ver que las potencias de Q tienden a O .

Definición 3.4 *Para una cadena de Markov absorbente definimos la **matriz fundamental** como $N = (I - Q)^{-1}$.*

Definición 3.5 *Definimos \mathbf{n}_j como una función cuyo valor es el número total de veces que el proceso está en el estado transitorio s_j .*

Sea \mathbf{T} el conjunto de estados transitorios de la cadena de Markov, si denotamos con $E_i[\mathbf{n}_j]$ al valor esperado de \mathbf{n}_j suponiendo que el proceso inicia en el estado s_i , tenemos el siguiente:

Teorema 3.3 [4] $\{E_i[\mathbf{n}_j]\} = N$.

Definición 3.6 *Sea \mathbf{t} una función cuyo valor está dado por el número de pasos (incluyendo la posición original) en los que el proceso se encuentra en un estado transitorio.*

Si el proceso inicia en un estado ergódico entonces $t = 0$. Si el proceso inicia en un estado transitorio, entonces t nos da el número total de pasos necesarios para alcanzar un estado ergódico. En una cadena absorbente éste es el *tiempo de absorción*.

Sea ξ un vector columna cuyas entradas son todas 1.

Teorema 3.4 [4] $\{E_i[t]\} = N\xi$.

4. Convergencia del Algoritmo Genético Elitista

Los Algoritmos Genéticos se usan a menudo para resolver problemas de optimización del tipo: $\max\{f(b)|b \in \mathbb{B}^l = \{0, 1\}^l\}$ suponiendo que $0 < f(b) < \infty$ para todo $b \in \mathbb{B}^l$ y $f(b) \neq \text{const}$. En esta sección se muestra el modelo desarrollado por Rudolph [7] mediante el cual se demuestra la convergencia del Algoritmo Genético Elitista (AGE). En [7] Rudolph modela el Algoritmo Genético Simple (AGS) (por AG Simple entendemos un AG que no usa elitismo) mediante una cadena de Markov finita homogénea. Cada estado i de la cadena de Markov corresponde a una posible población del AGS de tal manera que el espacio de estados es $\mathcal{S} = \mathbb{B}^{nl}$ donde n es el número de individuos de la población y l es la longitud de cada individuo. Definimos a $\pi_k^t(i)$ como el individuo k de la población i en el paso t . Dada la naturaleza del AGS, la matriz de transición \mathbf{P} que lo representa queda definida como $\mathbf{P} = \mathbf{CMS}$, donde \mathbf{C} , \mathbf{M} y \mathbf{S} son las matrices de transición de los operadores de cruza, mutación y selección respectivamente. Cuando se usa *mutación uniforme* los elementos de \mathbf{M} son $m_{ij} = p_m^{H_{ij}}(1 - p_m)^{N - H_{ij}} > 0$, en donde p_m es la probabilidad de mutación del AGS, H_{ij} es la distancia de Hamming entre los estados i y j , y $N = nl$. De lo anterior concluimos que \mathbf{M} es positiva. El uso de un operador de selección proporcional o de torneo [8] determina la existencia de una probabilidad estrictamente positiva de que la población quede intacta, lo cual asegura que los elementos de la diagonal s_{ii} de la matriz de transición del operador son positivos, por lo que se concluye que la matriz \mathbf{S} es columna-permisible. En resumen tenemos que la matriz \mathbf{M} es positiva y \mathbf{S} es columna-permisible. Luego, por el Lema 3.1, la matriz $\mathbf{P} = \mathbf{CMS}$ es positiva y por lo tanto primitiva. Para poder hablar de convergencia, a continuación se presenta la definición correspondiente para el AGS [7]:

Definición 4.1 Sea $Z_t = \max\{f(\pi_k^t(i))|k = 1, \dots, n\}$ una sucesión de variables aleatorias que denotan la mejor aptitud dentro de la población representada por el estado i en el paso t . Un algoritmo genético converge al óptimo global si y sólo si:

$$\lim_{t \rightarrow \infty} P\{Z_t = f^*\} = 1 \quad (1)$$

donde $f^* = \max\{f(b)|b \in \mathbb{B}^l\}$.

De esta manera entenderemos que el AGS converge al óptimo global de la función objetivo si la probabilidad de que éste se encuentre en la población tiende a 1 cuando el número de iteraciones tiende a infinito. En [7] Rudolph argumenta que en las aplicaciones del mundo real el AGS comúnmente mantiene a través del proceso evolutivo la mejor solución encontrada hasta el momento por lo que lo correcto es modelar el AGS

de tal manera. Así pues, consideraremos ahora agregar a la población del AGS un *súper individuo* que no tomará parte en el proceso evolutivo y que por facilidad en la notación será colocado en la primera posición a la izquierda, es decir, se podrá acceder a él mediante $\pi_0(i)$. Llamaremos a esta nueva versión Algoritmo Genético Elitista (AGE). El operador de elitismo estará representado por la matriz \mathbf{E} que lo que hará será actualizar un estado de tal manera que si éste contiene un individuo mejor que su actual *súper individuo* éste será reemplazado por aquél. En particular, sea:

$$i = (\pi_0(i), \pi_1(i), \pi_2(i), \dots, \pi_n(i)) \in \mathcal{S} \quad (2)$$

$\pi_0(i)$ es el *súper individuo* de la población (estado) i . Ahora bien, sea $b = \operatorname{argmax}\{f(\pi_k(i)) | k = 1, \dots, n\} \in \mathbb{B}^l$ el mejor individuo de la población i excluyendo el *súper individuo* y:

$$j \stackrel{\text{def}}{=} (b, \pi_1(i), \pi_2(i), \dots, \pi_n(i)) \in \mathcal{S} \quad (3)$$

entonces:

$$e_{ij} = \begin{cases} 1 & \text{si } f(\pi_0(i)) < f(b) \\ 0 & \text{de otra manera.} \end{cases} \quad (4)$$

La nueva matriz de transición para el AGE resulta del producto de una matriz que está compuesta por 2^l matrices \mathbf{P} , una por cada posible *súper individuo* acomodadas de manera que entre mejor sea su *súper individuo* más alta será su posición, y la matriz \mathbf{E} del operador de elitismo:

$$\begin{aligned} \mathbf{P}^+ &= \begin{pmatrix} \mathbf{P} & & & \\ & \mathbf{P} & & \\ & & \ddots & \\ & & & \mathbf{P} \end{pmatrix} \begin{pmatrix} \mathbf{E}_{11} & & & \\ \mathbf{E}_{21} & \mathbf{E}_{22} & & \\ \vdots & \vdots & \ddots & \\ \mathbf{E}_{2^l,1} & \mathbf{E}_{2^l,2} & \cdots & \mathbf{E}_{2^l,2^l} \end{pmatrix} \\ &= \begin{pmatrix} \mathbf{PE}_{11} & & & \\ \mathbf{PE}_{21} & \mathbf{PE}_{22} & & \\ \vdots & \vdots & \ddots & \\ \mathbf{PE}_{2^l,1} & \mathbf{PE}_{2^l,2} & \cdots & \mathbf{PE}_{2^l,2^l} \end{pmatrix} \end{aligned}$$

La estructura mostrada de la matriz \mathbf{P}^+ se debe a que, como ya se mencionó, las poblaciones están ordenadas de manera descendente de acuerdo a la calidad de su *súper individuo*. De tal manera, los espacios en blanco representan ceros puesto que no es posible pasar de un estado a otro con un *súper individuo* de menor calidad. De lo anterior se concluye que $\mathbf{PE}_{11} = \mathbf{P}$ puesto que tales matrices corresponden con las poblaciones que tienen como *súper individuo* al óptimo f^* .

Por otra parte, haciendo las siguientes definiciones:

$$\mathbf{R} = \begin{pmatrix} \mathbf{PE}_{21} \\ \vdots \\ \mathbf{PE}_{2^l,1} \end{pmatrix} \quad \mathbf{T} = \begin{pmatrix} \mathbf{PE}_{22} & & & \\ \vdots & \ddots & & \\ \mathbf{PE}_{2^l,2} & \cdots & \mathbf{PE}_{2^l,2^l} & \end{pmatrix}$$

concluimos que la matriz \mathbf{P}^+ es reducible a la forma:

$$\mathbf{P}^+ = \begin{pmatrix} \mathbf{P} & \mathbf{0} \\ \mathbf{R} & \mathbf{T} \end{pmatrix}. \quad (5)$$

Como conclusión tenemos el siguiente teorema:

Teorema 4.1 *El AGE converge al óptimo global.*

Demostración La submatriz \mathbf{P} contiene las probabilidades de transición de estados óptimos globales. Puesto que \mathbf{P} es una matriz estocástica primitiva y $\mathbf{R}, \mathbf{T} \neq 0$, el Teorema 3.1 garantiza que la probabilidad de permanecer en un estado no-óptimo converge a cero. Por lo tanto la probabilidad de permanecer en un estado óptimo global converge a 1. ■

5. Tiempo esperado de convergencia

En la sección 2 se presentaron algunos resultados acerca de la matriz fundamental correspondiente a la matriz de transición de una cadena de Markov. Como vimos, tal matriz puede usarse para calcular los tiempos esperados de convergencia de la cadena. En esta sección se aplicarán tales resultados a la correspondiente matriz de transición del AGE. Puesto que en la ecuación (5) la matriz \mathbf{P} corresponde con las poblaciones cuyo súper individuo es el óptimo global, podemos considerar que al estar la cadena en uno de esos estados, se ha terminado el proceso. Luego, cualquier cambio en la población puede ignorarse pues el súper individuo no se modificará. Así pues, podemos reescribir la matriz \mathbf{P}_+ como:

$$\mathbf{P}^+ = \begin{pmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{R} & \mathbf{T} \end{pmatrix}$$

Podemos ver claramente ahora que la cadena de Markov correspondiente al AGE es absorbente. De acuerdo a la Definición 3.4, la matriz fundamental que nos interesa en este caso es:

$$\mathbf{N} = (\mathbf{I} - \mathbf{T})^{-1}$$

Puesto que nuestro objetivo es conocer la matriz fundamental \mathbf{N} , procederemos a estudiar la estructura de la matriz bloque \mathbf{T} .

5.1. Matriz \mathbf{P}

En esta sección estudiaremos los elementos de la matriz \mathbf{P} . Como sabemos, dicha matriz es resultado del producto: $\mathbf{P} = \mathbf{C}\mathbf{M}\mathbf{S}$, por lo que a continuación se especificarán los elementos de las matrices correspondientes.

Elementos de la Matriz de Cruza. A continuación se presenta el modelo de la cruza uniforme. Para ello se definió la operación \oplus como la negación del *o-exclusivo*, y se aplicará entre los bits correspondientes a una posición fija de un padre y un posible hijo.

Cuando se lleva a cabo la cruce uniforme (suponiendo un porcentaje de cruce de 0.5) cada bit de un nuevo hijo tiene 0.5 de probabilidad de ser igual al correspondiente bit de cada uno de los dos padres. Usando este hecho se desarrolló la fórmula siguiente:

$$c_{ij} = \prod_{q=1}^n (0.5)^l \prod_{r=1}^l \sum_{s=\phi(q)-1}^{\phi(q)} \pi_s^r(i) \oplus \pi_q^r(j)$$

donde $\phi(q) = 2 \lfloor \frac{q+1}{2} \rfloor$.

Intuitivamente, la idea es que dados dos padres y un posible hijo, la probabilidad de obtener el segundo de los primeros no es más que el producto de la probabilidad de obtener de manera independiente cada bit del hijo a partir de los padres.

Elementos de la Matriz de Mutación. Como recordaremos, Rudolph modela el AG usando mutación uniforme. Los elementos correspondientes son:

$$m_{ij} = p_m^{H_{ij}} (1 - p_m)^{N - H_{ij}}$$

donde p_m es la probabilidad de mutación y H_{ij} es la distancia de Hamming entre las poblaciones i y j .

Elementos de la Matriz de Selección. La selección modelada es la selección proporcional. Como sabemos, en tal selección cada individuo tiene una probabilidad de ser seleccionado proporcionalmente a su aptitud, de donde:

$$s_{ij} = \begin{cases} \frac{\prod_{k=1}^n f(\pi_k(j))}{(\sum_{k=1}^n f(\pi_k(i)))^n} & \text{si } \pi_k(j) \in \{\pi_r(i) | r = 1, \dots, n\} \forall k = 1, \dots, n. \\ 0 & \text{de otra manera.} \end{cases}$$

en donde f es la función objetivo (aptitud).

Elementos de la Matriz P. Dado que la matriz \mathbf{P} es el producto de las matrices de cruce, mutación y selección, tenemos que:

$$p_{ij} = \sum_{p=1}^n \left(\sum_{q=1}^n c_{iq} m_{qp} \right) s_{pj}$$

por lo que:

$$p_{ij} = \sum_{p=1}^n \left[\sum_{q=1}^n \left(\prod_{r=1}^n (0.5)^l \prod_{s=1}^l \sum_{t=\phi(r)-1}^{\phi(r)} \pi_t^s(i) \oplus \pi_r^s(q) \right) (p_m^{H_{qp}} (1 - p_m)^{N - H_{qp}}) \right] \frac{\prod_{k=1}^n f(\pi_k(j))}{(\sum_{k=1}^n f(\pi_k(p)))^n}$$

5.2. Matriz E

A partir de las ecuaciones (2-4) se concluye que la matriz \mathbf{E} tiene exactamente un 1 por fila además de que sólo \mathbf{E}_{11} es una matriz identidad y las matrices \mathbf{E}_{aa} ($a \geq 2$) son matrices identidad con algunos ceros en la diagonal. Dada una población fija, es claro que su súper individuo no puede más que mejorar (figura 1).

Es claro también que la matriz \mathbf{E} tiene columnas de ceros: aquellas correspondientes a poblaciones cuyo súper individuo no sea el mejor individuo que contiene.

poblaciones que tienen como súper individuo al	óptimo	segundo mejor	...	peor
óptimo	E_{11}	0	0	0
segundo mejor	E_{21}	E_{22}	0	0
⋮	⋮	⋮	⋮	0
peor	$E_{2^l 1}$	$E_{2^l 2}$...	$E_{2^l j}$

Figura 1: La estructura de la matriz de elitismo se debe a que el súper individuo de una población dada sólo puede mejorar, por lo que existen bloques de ceros.

Asimismo, en aquellas columnas correspondientes a poblaciones cuyo súper individuo es el mejor de la población, tenemos que es posible encontrar más de un 1, de hecho, se encontrarán como máximo $(2^l - i + 1)$ unos por columna en el bloque de columnas correspondiente al i -ésimo mejor individuo (de los 2^l posibles). De esta manera, conforme descendemos a lo largo de las filas de bloques, las poblaciones se van “repartiendo” a lo largo de la fila dependiendo de la calidad de su máximo individuo. A partir de este momento asumiremos que dentro de cada bloque las poblaciones se encuentran ordenadas (por conjuntos dentro de los cuales no importa el orden) de acuerdo a la calidad de su máximo individuo.

5.3. Experimentos

Consideremos pues el caso más simple que se apega a las condiciones comunes de los AG. Sea $l = 2$ y $n = 2$, esto nos da un total de $2^{2 \times 2} = 2^4 = 16$ poblaciones a considerar por el AG. Se usó el paquete MATHEMATICA 4.0 para llevar a cabo los cálculos correspondientes a la matriz fundamental \mathbf{N} , en donde $\mathbf{N} = (\mathbf{I} - \mathbf{T})^{-1}$. Se llevaron a cabo las pruebas correspondientes al caso de 16 poblaciones y 4 individuos. En este caso cada población consta de 2 individuos y cada individuo es de longitud 2. Se definió la función: $f(x_1 x_2) = x_1 + 0.5x_2 + 0.5$, mediante la cual se asigna a cada individuo un valor distinto que permite ordenar (o jerarquizar) a la población, en función de este valor. Como la función f es estrictamente positiva para cualquier individuo, esta misma se usó como función de aptitud. En la siguiente tabla se muestran los cuatro posibles individuos, su aptitud correspondiente y su jerarquía (o posición dentro de la población):

individuo	aptitud	jerarquía
00	0.5	4
01	1.0	3
10	1.5	2
11	2.0	1

A continuación se presenta la tabla con las 16 poblaciones correspondientes organizadas de acuerdo al individuo máximo correspondiente:

Máximo	No.	<i>individuos</i>		Máximo	No.	<i>individuos</i>	
1	1	11	00	2	8	10	00
1	2	11	01	2	9	10	01
1	3	11	10	2	10	10	10
1	4	11	11	2	11	01	10
1	5	10	11	2	12	00	10
1	6	01	11	3	13	01	00
1	7	00	11	3	14	01	01
				3	15	00	01
				4	16	00	00

Como podemos ver, se tienen 7 poblaciones que tienen como máximo al óptimo, 5 que tienen como máximo al segundo mejor, 3 al tercero mejor y una al peor individuo. Además, dentro de estos grupos no importa el orden. De acuerdo al modelo descrito en la sección 3, se obtuvo la matriz fundamental \mathbf{N} . Puesto que en nuestro modelo la probabilidad de cruce (p_c) se fijó con el valor 1 y dado que el tamaño de la población y la longitud de los individuos están fijos también, los resultados sólo dependen de la probabilidad de mutación (p_m). Por otra parte, se corrió un AG con las condiciones impuestas por el modelo dentro de las que se hace notar que el súper-individuo (individuo elitista) no debe intervenir en el proceso evolutivo. Los parámetros fijos para el AG fueron:

$$\begin{aligned} \text{tamaño de población} &= 2 \\ \text{longitud del cromosoma} &= 2 \\ \text{probabilidad de cruce} &= 1.0 \end{aligned}$$

Sea g la variable aleatoria cuyo valor es el número de iteraciones necesarias para la convergencia del AG. A continuación se muestra el valor esperado de la variable g ($E[g]$) y la desviación estándar ($D[g]$) correspondiente, obtenidos (usando MATHEMATICA 4.0) por el modelo teórico (MT) desarrollado y los resultados obtenidos por el AG, para los cuales se hicieron 100 corridas con semilla de aleatorios distinta:

p_m	<i>AG</i>		<i>MT</i>	
	$E[g]$	$D[g]$	$E[g]$	$D[g]$
0.001	345.05	540.001	514.137	660.781
0.005	90.98	131.4615	103.782	132.47
0.01	40.89	54.01943	52.4913	66.4335
0.03	13.0	16.8367	18.3079	22.4155
0.07	5.08	7.036241	8.56138	9.85261
0.1	4.69	6.5	6.38044	7.03
0.2	2.69	3.47	3.87256	3.78054
0.5	1.32	1.847083	2.52747	1.96485

En la figura 2 se muestra la gráfica de los valores obtenidos por ambos métodos.

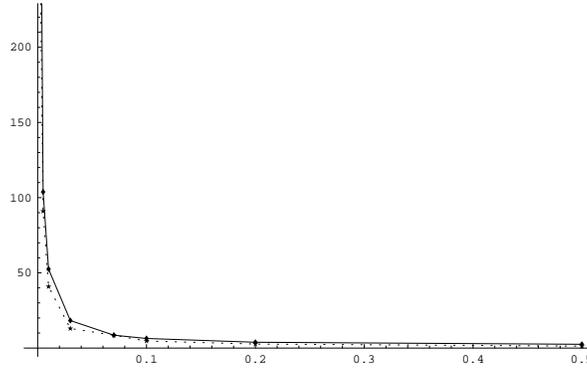


Figura 2: Gráfica de los valores obtenidos por el AG y por el modelo teórico desarrollado para la variable g , en función de la probabilidad de mutación. La línea continua corresponde con los valores generados por el modelo teórico y la línea punteada con los valores generados por el AG.

Tal vez los valores que se obtuvieron en los casos de $p_m = 0.001, 0.005, 0.01$ parecerán muy altos, pero son correctos desde el siguiente punto de vista: si recordamos la tabla de todas las posibles poblaciones, aproximadamente el 46 % (poblaciones 1 a 7) contiene al óptimo, y sólo las poblaciones 9 y 11 contienen los esquemas necesarios para dar lugar al óptimo mediante la cruce. En conclusión, aproximadamente el 40 % de las posibles poblaciones no contiene al óptimo ni a los esquemas necesarios para dar lugar a él. Dado que la población es muy pequeña el número de mutaciones llevadas a cabo en un número corto de generaciones es prácticamente cero, siendo éste el único mecanismo para poder converger correctamente, por lo que se necesita de un número alto de generaciones para encontrar finalmente el óptimo. En la siguiente tabla se muestran las poblaciones para las cuales es particularmente difícil alcanzar el óptimo. Recuérdese que el súper-individuo no participa en el proceso evolutivo.

Máximo	No.	<i>individuos</i>		Máximo	No.	<i>individuos</i>	
2	8	10	00	3	13	01	00
2	10	10	10	3	14	01	01
2	12	00	10	3	15	00	01
				4	16	00	00

6. Conclusiones

Los modelos existentes para estimar el tiempo de convergencia del AG son bastante burdos y, por la misma razón, bastante lejanos al comportamiento real de dicho algoritmo. Se desarrolló un mecanismo para estimar el tiempo de convergencia del AG mediante un modelo basado en cadenas de Markov. Los resultados de los experimentos llevados a cabo para el caso más sencillo considerado *real* nos llevan a concluir que el modelo es correcto. Sin embargo, está claro que en tal caso fue relativamente

sencillo obtener la matriz \mathbf{P} correspondiente, proceso que por lo general será cada vez más complicado conforme el tamaño de la población y de los individuos aumenta. Supongamos que la población consta de 50 individuos de longitud 5, la matriz \mathbf{P} correspondiente es de tamaño $2^{250} \times 2^{250}$. En general, el tamaño de la matriz crece de manera exponencial. Así pues, podemos concluir que, desde un punto de vista práctico, las cadenas de Markov no resultan una herramienta teórica recomendable para este análisis. Sin embargo, existen otras alternativas a las que se podría recurrir, como por ejemplo: la mecánica estadística [6], los enfoques de interpretación geométrica [9] y el modelado con base en una búsqueda aleatoria y acoplamiento variable de los operadores genéticos [10].

Agradecimientos

La primera autora agradece a CONACyT por la beca otorgada para realizar estudios de postgrado en la Sección de Computación del Departamento de Ingeniería Eléctrica del CINVESTAV-IPN. El segundo autor agradece el apoyo proporcionado por CONACyT a través del proyecto 34201-A.

Referencias

- [1] Carol A. Ankenbrandt. An extension to the theory of convergence and a proof of the time complexity of genetic algorithms. In Gregory J. E. Rawlins, editor, *Foundations of Genetic Algorithms*, pages 53–68. Morgan Kaufmann Publishers, 1991.
- [2] D. E. Goldberg and P. Segrest. Finite markov chain analysis of genetic algorithms. In *Proceedings of the Second International Conference on Genetic Algorithms*, pages 1–8. Lawrence Erlbaum Associates, 1987.
- [3] M. Iosifescu. *Finite Markov Processes and Their Applications*. Wiley, Chichester, 1980.
- [4] John G. Kemeny and J. Laurie Snell. *Finite Markov Chains*. D. Van Nostrand Company, Inc., Princeton, New Jersey, 1960.
- [5] Sushil J. Louis and Gregory J.E. Rawlins. Syntactic analysis of convergence in genetic algorithms. In L. Darrell Whitley, editor, *Foundations of Genetic Algorithms*, pages 141–151, San Mateo, CA, 1993. Morgan Kaufmann Publishers.
- [6] A. Prügel-Bennett and J.L. Shapiro. An analysis of genetic algorithms using statistical mechanics. *Physical Review Letters*, 72(9):1305–1309, 1994.
- [7] Günter Rudolph. Convergence properties of canonical genetic algorithms. *IEEE Transactions on Neural Networks*, 1(5):96–101, 1994.
- [8] Günter Rudolph and Alexandru Agapie. Convergence properties of some multi-objective evolutionary algorithms. In *Proceedings of the 2000 Conference on Evolutionary Computation*, pages 1010–1016, Piscataway, New Jersey, 2000. IEEE.
- [9] Michael D. Vose. Modeling simple genetic algorithms. In L. D. Whitley, editor, *Foundations of Genetic Algorithms 2*, pages 63–73, San Mateo California, 1991. Morgan Kaufmann Publishers.
- [10] Michael D. Vose. *The Simple Genetic Algorithm: Foundations and Theory*. MIT Press, Cambridge, Massachusetts, 1999.