

Un Estudio Empírico sobre el Impacto de la Mutación en la Variedad Genética y la Rata de Convergencia de un Algoritmo Genético Simple

Sandra Lima¹, Minoru Akiyama²
slima@ucla.edu.ve, minorucho@gmail.com

¹ Unidad de Inteligencia Artificial, Universidad Centroccidental Lisandro Alvarado, Barquisimeto, Venezuela

² Departamento de Matemática, Universidad Centroccidental Lisandro Alvarado, Barquisimeto, Venezuela

Resumen: Los Algoritmos Genéticos Simples (AGS), son métodos estocásticos de búsqueda que emulan los principios básicos de la evolución biológica. La dinámica del AGS se centra en los cambios realizados, mediante selección, cruce y mutación, a los genotipos de una población de individuos, los cuales representan posibles soluciones al problema planteado. En consecuencia, estos operadores deben ser cuidadosamente diseñados para mantener un equilibrio entre la variedad genética y la rata de convergencia poblacional. Esta investigación plantea un estudio empírico sobre algunos operadores de mutación no adaptativos y analiza sus efectos en la población, en cuanto a la variedad genética y la rata de convergencia medida en épocas. Se presenta una novedosa pseudo-métrica de Distancia Poblacional entre Generaciones, cuyo método es flexible y dinámico, para evaluar la evolución temporal de la variedad entre poblaciones genéticas. Se excluye el operador de cruce en el proceso evolutivo para estudiar solo el impacto de la mutación en la convergencia y en la variedad poblacional. Luego se experimenta con el algoritmo genético completo para establecer comparaciones entre los resultados obtenidos.

Palabras Clave: Operador Genético; Mutación; Algoritmo Genético Simple; Variedad Genética; Convergencia Poblacional; Distancia Genética.

Abstract: Simple Genetic Algorithms (SGA), are stochastic search methods that emulate the basic principles of biological evolution. Dynamics of SGA focuses on changes made by selection, crossover and mutation, to the genotypes of a population of individuals which represent possible solutions to the problem. Consequently, these operators must be carefully designed to maintain a balance between genetic variation and population convergence. This research presents an empirical study of some non-adaptive mutation operators and analyzes their effects on the population, in terms of genetic variation and rate of convergence measured by epochs. A novel pseudo-metric population distance between generations is presented, whose method is flexible and dynamic, for the evaluation of the evolution of genetic variation among populations. The crossover operator is excluded of the evolutionary process to only study the impact of the mutation on the population convergence and population variety. Then we experiment with the complete genetic algorithm to establish comparisons between the obtained results.

Keywords: Genetic Operator; Mutation; Simple Genetic Algorithm; Genetic Diversity; Population Convergence; Genetic Distance.

I. INTRODUCCIÓN

Los Algoritmos Genéticos Simples (AGS), propuestos en la década de los 60 por John Holland y sus estudiantes [1] mientras buscaban simular sistemas adaptativos biológicos en medios artificiales, representan hoy en día uno de los métodos de computación evolutiva más usados y explorados en la literatura especializada. La dinámica del AGS se centra en los cambios realizados a los genotipos de una población de individuos que representan cada uno una posible solución al problema planteado. Estos cambios se aplican de manera repetitiva a través de los llamados operadores evolutivos o de

cambio como el cruce, la mutación y la selección, creando ciclos conocidos como generaciones o épocas evolutivas.

Debido a que el AGS opera sobre los genotipos de una población, es importante mantener un equilibrio entre la variedad genética y la rata de convergencia poblacional. Esto implica que para que el AGS sea eficiente (y eficaz), los operadores de cambio deben ser cuidadosamente diseñados y/o configurados para evitar los fenómenos conocidos como desviación genética (genetic drift) y convergencia prematura; el primero ocasiona que todo el banco genético de la población avance hacia un mismo patrón (individuo) y el segundo, que la convergencia se 'estaque' en pocas generaciones o épocas

evolutivas, posiblemente en algún óptimo local. Esta investigación plantea un estudio empírico sobre algunos operadores de mutación no adaptativos y analiza sus efectos en la población considerando la variedad genética y la tasa de convergencia medida en épocas.

II. EL OPERADOR DE MUTACIÓN COMO RESPONSABLE DE LA VARIEDAD GENÉTICA

El operador de mutación es la principal herramienta evolutiva en un AGS. Si bien el cruce con recombinación introduce capacidades de adaptación ante un ambiente específico, la mutación es responsable de introducir nuevo material genético en los alelos que serán transmitidos a las nuevas generaciones. Esto puede resultar en un impacto positivo si el fenotipo presenta una mayor adaptación al contexto evaluado, negativo si el resultado produce un menor nivel de adaptación, o neutro si los cambios en el genotipo no alteran el fenotipo y por ende el nivel de adaptación permanece igual. En cualquier caso, es la herramienta necesaria para evitar caer en óptimos locales cuando la diversidad genética del grupo carece del potencial para hacerlo.

Sin embargo, la mutación frecuentemente es considerada como un operador secundario de poca importancia, debido a que tiende a distraer al algoritmo durante la búsqueda de convergencia [2]. Si la tasa de mutación es grande, la diversidad genética resultante ubica los fenotipos de los individuos en regiones dispersas en el espacio solución y por tanto obstruye la convergencia, mientras que una tasa muy pequeña resultará en una exploración muy minuciosa de una región específica y por ende demora la convergencia con la debilidad de caer en la atracción de un óptimo local. Por ello es interesante plantearse cuál sería la tasa de mutación ideal para un tipo de problema dado y la variabilidad que dicha tasa de mutación le introduce a una población genética.

Debido a la naturaleza estocástica de los operadores del AGS y sus relaciones, no es posible construir un modelo determinista y en consecuencia, predictivo del comportamiento del algoritmo frente a una determinada configuración de parámetros y un problema de una naturaleza dada. Sin embargo, los ensayos y pruebas empíricas proporcionan una firme base estadística para comprender mejor estos operadores y la mecánica del AGS, ayudando así al diseño eficiente, formulación y escogencia de parámetros para un problema específico. Se plantea entonces este trabajo de investigación de campo, donde se aísla el efecto de la mutación en la variedad genética y convergencia poblacional de un AGC, para una vez aplicadas las diferentes métricas y los exámenes estadísticos de rigor, formular planteamientos de discusión del desempeño y papel de la mutación en el camino evolutivo.

III. DISEÑO DEL AGS

En las investigaciones de Fogarty y De Jong [3][4] se usó un operador de mutación basado en el original propuesto por Holland en su Algoritmo Genético Simple [1], donde los genes mutantes son escogidos de manera aleatoria mediante un experimento de Bernoulli evaluando todo el cromosoma.

En este caso el AGS está constituido por una población de “n” individuos. Cada individuo de la población se encuentra

representado por un cromosoma binario en código Gray de longitud “m”. La ventaja de utilizar código gray es que mantiene una distancia Hamming entre los cromosomas de manera que pequeños cambios en el genotipo introduzca pequeños cambios en el fenotipo y viceversa [5]. La longitud del cromosoma se divide en partes iguales de acuerdo al número de variables de la función que se desea optimizar; cada segmento del cromosoma representa a una variable en el fenotipo.

El algoritmo funciona como sigue: se decodifica, de Gray a real normalizado, el segmento correspondiente a cada variable. Luego se calcula la función objetivo para asignar el fitness o valor de adaptación al individuo. Los valores fitness de los individuos son utilizados por el operador de selección para elegir dos padres que generan dos hijos en la nueva generación. Está configurado para trabajar estilo “ruleta rusa” [1]. Luego se aplica el operador de cruce en un punto para cada uno de los segmentos [1]. Este paso se omite para los experimentos que aíslan el operador de cruce. Finalmente, se aplica mutación en los bits o alelos del cromosoma seleccionado, de acuerdo a configuraciones que serán detalladas en la siguiente sección. Este proceso se repite por un máximo de generaciones “maxgen” dado.

IV. OPERADORES DE MUTACIÓN NO ADAPTATIVOS

Dos aspectos importantes en la implementación de un operador de mutación para un AGS, son el tipo, modo o forma como se aplicará el cambio al genotipo, y la tasa o frecuencia estadística con la que ocurrirán estos cambios [6]. En este sentido, algunas investigaciones empíricas han sido desarrolladas para profundizar en los efectos de este operador; Fogarty estudió 5 esquemas de tasa de mutación variable [7] en un modelo de quemador industrial.

De Jong [4] mostró que con un tamaño de población pequeño y una alta tasa de mutación, el desempeño del algoritmo mejora durante las primeras generaciones utilizando una métrica que considera la convergencia de la población. Bäck y Schütz [2] [8] en un análisis más reciente argumentaron que para un tipo de problema específico, la tasa de mutación óptima es $1/N$ donde N es el número de bits del cromosoma. Ochoa, Harvey y Buxton [9] obtuvieron el umbral del error a partir del cual la tasa de mutación produce peores soluciones y su incidencia en la presión selectiva.

A continuación se diseñan e implementan varios operadores de mutación no adaptativos, los cuales se diferencian por la cantidad de ruido o cambio que introducen en los cromosomas. Los individuos elegidos para mutar son determinados aleatoriamente por una probabilidad fija de mutación. El cambio introducido en un cromosoma se aplica a segmentos de longitud variable y se cambian los alelos dentro de ese segmento mediante una variable de Bernoulli, la cual retorna un valor booleano dependiendo de la proporción de ruido dada como parámetro.

A. Mutación Clásica

El operador de mutación se aplica a todos los individuos seleccionados para la nueva generación, en alelos escogidos aleatoriamente por la proporción de ruido. A nivel fenotípico, podría afectar a todas las variables de la función objetivo. La inclusión de ruido en un cromosoma está directamente relacionada con la variable proporción de ruido.

B. Muta Uno

En cada cromosoma, muta un único alelo elegido al azar. Se considera el operador con menor inclusión de ruido.

C. Mutan Segmentos Aleatorios

Dentro de cada cromosoma, se aplica ruido a un segmento de longitud aleatoria. Este segmento puede incluir, a nivel fenotípico, una o varias variables de la función objetivo.

D. Muta Variable

Se elige al azar una variable del fenotipo por cada individuo de la población. En el cromosoma, se aplica ruido al segmento que representa dicha variable.

V. MÉTRICAS DE EVALUACIÓN PARA LA CONVERGENCIA Y VARIEDAD POBLACIONAL

A. Rata de Convergencia

La convergencia poblacional se medirá en términos del fitness promedio por generación a lo largo de todos los experimentos. Como el promedio se ve afectado por valores extremos encontrados por el proceso de mutación, también se calculará la mediana poblacional como el valor central que reúne el 50% de los individuos y fija la tendencia de la población hacia el óptimo. Además, se considerará el valor mínimo de mediana por experimento que con un error máximo dado, reúna al menos al 25% del total de generaciones.

B. Variedad Poblacional

1) *Distancia Hamming entre Vecinos*: Se calculará la métrica de Distancia Haming Promedio entre vecinos de la población genética para estudiar la variedad poblacional. La relación entre vecinos es circular, es decir, se asume que el vecino del último individuo es el primero de la población.

2) *Cantidad Promedio de Individuos Diferentes*: La cantidad de individuos diferentes por generación puede describir la tendencia de la población hacia la convergencia. Un individuo se diferencia de otro en al menos un alelo.

3) *Distancia Poblacional entre Generaciones*: Se plantea una medida poblacional basada en distancia entre conjuntos en un espacio métrico [10], donde los elementos son los individuos de la población actual, representados en su forma binaria como la matriz A, la cual es comparada con la población de la generación anterior denotada como la matriz B. Cuando la distancia es menor más parecidas son las poblaciones y viceversa.

Por métrica se tiene una función $d: \zeta_m \times \zeta_m \rightarrow R$ que cumple las siguientes propiedades:

- $\forall a, b \in \zeta_m, d(a, b) \geq 0$
- $\forall a, b \in \zeta_m, d(a, b) = 0 \leftrightarrow a = b$
- $\forall a, b \in \zeta_m, d(a, b) = d(b, a)$
- $\forall a, b, c \in \zeta_m, d(a, c) \leq d(a, b) + d(b, c)$

Donde ζ_m representa el espacio de cromosomas. Algunas métricas que se pueden usar en ζ_m son las siguientes:

- *Métrica de Hamming Normalizada*

$$d_H(a, b) = \frac{1}{m} \sum_{j=0}^{m-1} \delta(a_j, b_j), \forall a, b \in \zeta_m$$

$$\text{Donde, } \delta(a_j, b_j) = \begin{cases} 0 & \text{si } a_j = b_j \\ 1 & \text{si } a_j \neq b_j \end{cases},$$

$$a = a_0, \dots, a_{m-1}; b = b_0, \dots, b_{m-1} \quad (1)$$

- *Métrica Trivial*

$$d_T(a, b) = \begin{cases} 0 & \text{si } a = b \\ 1 & \text{si } a \neq b \end{cases}, \forall a, b \in \zeta_m \quad (2)$$

- *Métrica de Cantor Normalizada*

$$d_C(a, b) = \frac{1}{2 \times (1 - \frac{1}{2^m})} \sum_{j=0}^{m-1} \frac{\delta(a_j, b_j)}{2^j}, \quad \forall a, b \in \zeta_m \quad (3)$$

Las métricas que aportan más al estudio de la variedad genética a lo largo de las generaciones son las métricas de Cantor y de Hamming. La métrica de Cantor introduce una variación en la distancia de Hamming en cuanto a que le da menor peso a las diferencias de genes a medida que estos ocupan una posición más a la derecha de los cromosomas.

En el caso de funciones objetivo multivariables cada cromosoma del AGS está compuesto por un número finito de subcromosomas $a = a_{(1)}, a_{(2)}, \dots, a_{(l)}$ concatenados en ese orden. En este caso se denomina $\zeta_m = \zeta_{m_1+m_2+\dots+m_l}$ a la colección de todos los cromosomas de este tipo para hacer énfasis en que un cromosoma esta subdividido en l subcromosomas, el primero de longitud m_1 , el segundo de longitud m_2 y así sucesivamente. En este trabajo, los subcromosomas tienen igual longitud.

Considerando esta concatenación se puede diseñar nuevas medidas de distancia para el cromosoma a partir de la combinación de diferentes métricas para cada segmento, según las características del espacio solución. Una forma sencilla de calcular la distancia global d_g entre dos cromosomas sería calculando un promedio de distancias entre sus segmentos:

$$d_g(a, b) = \frac{d_x(a_{(1)}, b_{(1)}) + \dots + d_x(a_{(l)}, b_{(l)})}{l} \quad (4)$$

Donde d_x es la métrica de distancia individual elegida para cada segmento. También se podría trabajar en función de los máximos o mínimos entre las distancias individuales. La función d_g es una métrica y el par (ζ_m, d_g) es un espacio métrico completo y compacto. Llamaremos al espacio (ζ_m, d_g) espacio de g-cromosomas.

a) Población de Cromosomas

Se tiene una población de n individuos y a cada individuo del AGS le corresponde un cromosoma de longitud m . Esta población se representa por una matriz $A_{n \times m}$ donde cada componente o alelo es cero (0) o uno (1).

Denotemos por $M_{n \times m}$, al conjunto de todas las matrices de tamaño $n \times m$ compuestas por cromosomas en ζ_m . El interés se centra en construir una métrica adecuada que describa cuando dos poblaciones $A, B \in M_{n \times m}$ son cercanas o distantes. Tal métrica debe tomar en cuenta las siguientes consideraciones:

- La distancia entre A y B debe ser definida comparando los individuos, es decir, las filas de A con las filas de B.

- Debe tomarse en cuenta que los individuos de una población pueden cambiar de lugar sin alterar a la población. Por ejemplo, $A = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \\ 1 & 1 & 0 & 0 \end{bmatrix}$ y $B = \begin{bmatrix} 0 & 0 & 0 & 1 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$ son dos matrices muy distintas que representan a la misma población. Por tanto, la métrica debe tomar en cuenta el orden de las filas de una matriz.
- El número de veces que cada cromosoma se repite en una población debe ser considerado. No es lo mismo la población dada por:

$$A = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \end{bmatrix}, \text{ que la dada por: } B = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \end{bmatrix}$$

Un censo poblacional permite establecer estas diferencias entre poblaciones [10]. Si $A \in M_{n \times m}$ representa una población, se extiende dicha matriz de manera que a cada fila $A^{(r)}$ se le agrega un nuevo subcromosoma $A_{(l+1)}^{(r)}$ de longitud m_{l+1} , siendo l la cantidad de subcromosomas que representan el problema a resolver. Este subcromosoma puede representar en código binario un valor según los siguientes criterios:

- El número de veces que la fila $A^{(r)}$ se encuentra repetida en la matriz A.
- El número de veces que la fila $A^{(r)}$ se encuentra repetida en la matriz B.

El valor resultado se encuentra acotado en $[0, n]$. En consecuencia, la longitud m_{l+1} debe ser escogida de forma tal que permita su representación en binario. Utilizando el ejemplo anterior, con $m_{l+1} = 3$, al aplicar el censo según el criterio 'i' las matrices A y B se convierten en:

$$A_c = \begin{bmatrix} 1 & 1 & 0 & 0 & | & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & | & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & | & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & | & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & | & 1 & 0 & 0 \end{bmatrix} \text{ y } B_c = \begin{bmatrix} 1 & 1 & 0 & 0 & | & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 & | & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & | & 0 & 0 & 1 \\ 1 & 1 & 0 & 0 & | & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 & | & 1 & 0 & 0 \end{bmatrix}$$

Aplicando el censo, A_c y B_c contienen una mejor información sobre la población que las dadas por las matrices A y B. Igualmente en las poblaciones censadas, la distancia entre individuos puede calcularse siguiendo las métricas antes descritas.

b) Pseudo-Métrica de Distancia Poblacional entre Generaciones

Sean las matrices $A, B \in M_{n \times m}$ y sus respectivas matrices censadas A_c y B_c . Llamaremos distancia dirigida de A hacia B, al valor:

$$\vec{D}(A, B) = \max_{1 \leq r \leq n} (\min_{1 \leq s \leq n} d_g(A_c^{(r)}, B_c^{(s)})) \quad (5)$$

Se observa que para r fijo, $\min_{1 \leq s \leq n} d_g(A_c^{(r)}, B_c^{(s)})$ mide la distancia que hay entre el cromosoma del individuo r en la población A con el individuo de la población B que más se le parece. Así, $\vec{D}(A, B)$ determina que tan diferente es la población de A con respecto a B. Esta relación no es

conmutativa ya que no se cumple que los individuos más cercanos a la población A, sean los mismos cuando se calcula \vec{D} en la población B. Formalmente se define la pseudo-métrica de distancia poblacional $d: M_{n \times m} \times M_{n \times m} \rightarrow \mathbb{R}$ como la función dada por:

$$d(A, B) = d(B, A) = \max\{\vec{D}(A, B), \vec{D}(B, A)\} \quad (6)$$

$$\forall A, B \in M_{n \times m}$$

Es una pseudo-métrica ya que las poblaciones A y B pueden ser distintas y sin embargo $d(A, B) = 0$. Aun en este caso, aunque $A \neq B$ las poblaciones que ellas representan son exactamente las mismas, solo habría que permutar las filas de A para obtener las filas de B. Dos individuos serán cercanos si son parecidos en el genotipo y si se repiten de forma semejante en la población censada.

Se aplicará esta métrica de distancia a las poblaciones formadas por la generación anterior y la nueva a lo largo del proceso evolutivo.

VI. EXPERIMENTACIÓN

Los experimentos comienzan con una población inicializada aleatoriamente. Se realizan 05 experimentos de 400 generaciones cada uno para una población de 30 individuos. Los individuos están compuestos por un cromosoma de longitud 32 en un contexto de minimización. Adicionalmente, se incluye un sub-cromosoma Gray de longitud 07 para el cálculo de las matrices censadas según el criterio 'i'. Los parámetros restantes son decididos en tiempo de ejecución. Entre ellos están: el tipo de mutación, acotamiento de valor para las variables de la función objetivo a optimizar, probabilidad de cruce, probabilidad de mutación y proporción de ruido introducido en los cromosomas mutados. Una probabilidad de cruce con valor 0 es utilizado para excluir dicho operador del proceso evolutivo y estudiar únicamente el impacto de la mutación en la convergencia y en la variedad poblacional. También se utiliza el algoritmo genético completo para establecer comparaciones entre los resultados obtenidos.

Para el cálculo de la distancia poblacional entre generaciones, se aplica la métrica de Cantor normalizada (3) en cada uno de los segmentos del cromosoma y la distancia entre individuos viene representada por el promedio resultante. En el caso de la convergencia, además del fitness y mediana promedio por generación, en cada experimento se cuenta el valor mínimo de la mediana que con un error máximo de 0.001 reúna al menos al 25% del total de generaciones. En este caso se podría concluir que el algoritmo ha convergido a dicho valor. Se eligieron 03 funciones bidimensionales en un contexto de minimización:

- $f_1(x, y) = (x - 2)^2 + (y - 2)^2 + 1$. Esta función objetivo tiene forma parabólica y tiene un mínimo conocido para $x = 2, y = 2, f_1(x, y) = 1$.
- $f_2(x, y) = x^2 \text{Sen}(x) + y^2 \text{Sen}(y) + 4$. Es una función multimodal con varios puntos máximos. Alcanza su mínimo en $f_2(x, y) = -44,53$.
- $f_3(x, y) = (xy(1 - x)(1 - y)\sin(9\pi x)\sin(9\pi y))^2$. Esta función muestra una malla interesante donde hay

muchos valores de (x,y) que minimizan la función en cero (0).

Se consideraron 04 casos experimentales donde se aplican los 04 operadores de mutación no adaptativos descritos en la sección IV:

- a) Baja probabilidad de mutación con una conservadora inclusión de ruido en los cromosomas mutados.
- b) Baja probabilidad de mutación con una alta inclusión de ruido en los cromosomas mutados.
- c) Alta tasa de mutación con una conservadora inclusión de ruido en los cromosomas mutados.
- d) Alta tasa de mutación con alta inclusión de ruido en los cromosomas mutados.

Para referirse a un experimento específico se tomará por ejemplo la notación '1a' para indicar que corresponde a la función objetivo Nro. 1 en el caso a. A continuación las discusiones de cada caso.

A. Caso a

Se probó $f_1(x,y)$ con $(x,y) \in [-2,6]$, $p_{cruce} = 0.8$, $p_{mutac} = 0.05$, $ruidomutac = 0.3$.

Los resultados indican que no hay una diferencia marcada entre aplicar el algoritmo genético completo y aislar el operador de mutación ya que la tasa es muy pequeña y el ruido introducido también lo es. La distancia poblacional muestra cambios de distancia de una generación a otra con una amplitud aproximada menor o igual a 0.1. La variabilidad entre individuos es relativamente constante.

En cuanto a la distancia dirigida, la población final 'F' se parece más a la inicial 'I' que al contrario y esto se corrobora a lo largo de los todos casos tanto si se aísla la mutación como si se considera el AGS completo. Es un resultado interesante ya que a nivel biológico esto consecuencia natural del proceso evolutivo (Figura 1). Por los resultados de la mediana se considera que el algoritmo ha convergido y que el cruce beneficia la variabilidad por encima de la mutación.

B. Caso b

Se probó f_1 con $(x,y) \in [-2,6]$, f_2 con $(x,y) \in [-6,8]$, $p_{cruce}=0.8$, $p_{mutac}=0.05$, $ruidomutac=0.6$ y f_3 con $(x,y) \in [-6,6]$, $p_{cruce}=0.8$, $p_{mutac}=0.05$, $ruidomutac=0.7$. Para el caso 1b se observa que hay mayor variabilidad en el cálculo del fitness y mediana promedio con respecto al caso a (Figuras 2 y 3). Esto se debe a que los individuos elegidos para mutar contienen mayor cantidad de ruido. El operador 'muta variable' se encuentra alejado de la tendencia global hacia el óptimo y los que más se acercaron fueron 'muta uno' y 'segmentos aleatorios'. El cálculo de la mediana demuestra la convergencia del algoritmo con resultados similares al caso 1a (Tabla I).

En el caso 2b se observan resultados muy similares al caso 1b en relación a la tendencia de los operadores. En cuanto al cálculo de la mediana los operadores que muestran mayor tendencia al óptimo fueron 'segmentos aleatorios', 'muta uno' y 'mutación clásica', aunque no se alcanza la convergencia en todos los experimentos (Tabla II). Considerando solo mutación, el comportamiento de cada operador, a pesar de no ser el ideal, se mantiene cercano al mínimo global. El

operador 'muta variable' continúa alejado de la tendencia global hacia el óptimo (Figura 4).

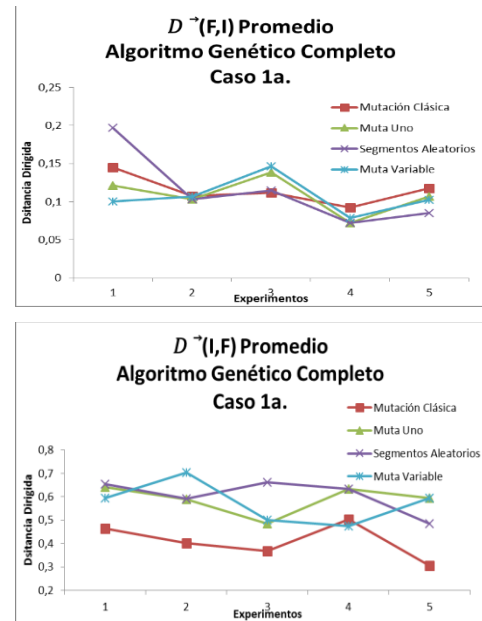


Figura 1: Distancia Dirigida Promedio entre las Poblaciones I y F Caso 1a

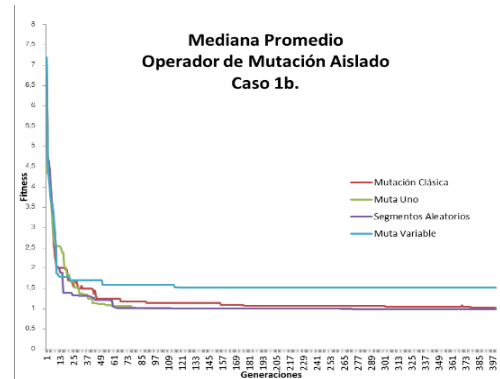


Figura 2: Mediana Promedio - Caso 1b

Tabla I: Mediana Mínima - Caso 1b

| Exp | Mutación Clásica | Muta Uno | Segmentos Aleatorios | Muta Variable |
|-----|------------------|----------|----------------------|---------------|
| 1 | 1,00194 | 1,000067 | 1,000005 | 1,065322 |
| 2 | 1,000618 | 1,000026 | 1,000391 | 2,137649 |
| 3 | 1,001181 | 1,000166 | 1,001044 | 1,083951 |
| 4 | 1,008906 | 1,000593 | 1,0004 | 1,036116 |
| 5 | 1,004692 | 1,000286 | 1,000093 | 1,077673 |

En caso 3b la mediana refleja convergencia del algoritmo, donde la mutación aislada obtuvo mejores resultados (Tablas III y IV). Sin embargo, por ser una función de superficie oscilante, el fitness promedio refleja fluctuaciones importantes en los operadores 'mutación clásica' y 'segmentos aleatorios' (Figura 5).

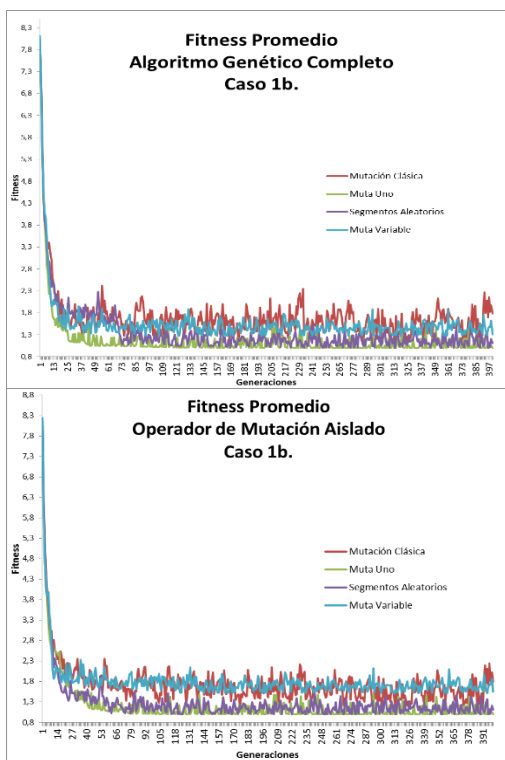


Figura 3: Fitness Promedio - Caso 1b

Tabla II: Mediana Mínima - Caso 2b

| Exp | Mutación Clásica | Muta Uno | Segmentos Aleatorios | Muta Variable |
|-----|------------------|-----------|----------------------|---------------|
| 1 | -43,599379 | -44,16573 | -43,520587 | -43,488156 |
| 2 | -43,869805 | -44,16526 | -44,165029 | -38,996075 |
| 3 | -43,816227 | -44,16496 | -44,165046 | -42,113712 |
| 4 | -43,845938 | -44,16576 | -44,165408 | -36,265633 |
| 5 | -44,043416 | -44,16523 | -24,028261 | -43,700752 |

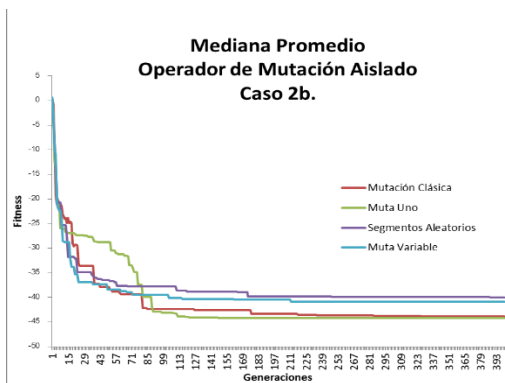


Figura 4: Mediana Promedio - Caso 2b

Tabla III: Mediana Mínima: AGS - Caso 3b

| Exp | Mutación Clásica | Muta Uno | Segmentos Aleatorios | Muta Variable |
|-----|------------------|----------|----------------------|---------------|
| 1 | 0,000588 | 0,000635 | 0,000077 | 0,000538 |
| 2 | 0,000761 | 0,00054 | 0,000177 | 0,000708 |
| 3 | 0,000245 | 0,000552 | 0,000374 | 0,000393 |
| 4 | 0,000852 | 0,000922 | 0,000757 | 0,00051 |
| 5 | 0,000395 | 0,000151 | 0,00031 | 0,000016 |

Tabla IV: Mediana Mínima: Operador de Mutación Aislado - Caso 3b

| Exp | Mutación Clásica | Muta Uno | Segmentos Aleatorios | Muta Variable |
|-----|------------------|----------|----------------------|---------------|
| 1 | 0,000357 | 0,000358 | 0,000068 | 0,000267 |
| 2 | 0,000015 | 0,000011 | 0,000146 | 0,000015 |
| 3 | 0,000342 | 0,000342 | 0,000342 | 0,000342 |
| 4 | 0,000026 | 0,000207 | 0,000111 | 0,000821 |
| 5 | 0,000004 | 0,000807 | 0,000081 | 0,000594 |

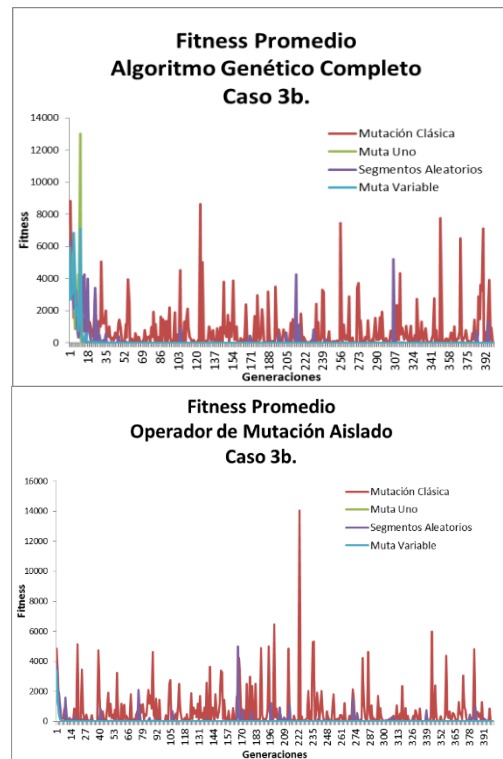


Figura 5: Fitness Promedio - Caso 3b

En la Figura 6 la distancia poblacional aumentó su amplitud en 0.03 aproximadamente con respecto al caso 1a. En caso de 2b la dispersión se mantuvo de forma similar a 1b para los distintos operadores. En el caso 3b la dispersión fue mayor por el aumento del ruido en los cromosomas mutados. Esto indica que a nivel genético la variedad se mantiene pese al cambio de función objetivo. Para el caso b, como lo muestra por ejemplo la Figura 7, la tendencia evolutiva de la distancia dirigida entre poblaciones 'I' y 'F' es similar a la observada en el caso a.

La distancia Hamming promedio entre vecinos de la Figura 8 muestra mayores fluctuaciones que el caso a por el ruido introducido en los cromosomas mutados. Del mismo modo, en el caso 2b y 3b la distancia Hamming tiende a disminuir a medida que transcurren las generaciones, siendo los operadores 'muta uno' y 'muta variable' los de menor distancia entre vecinos.

En la Figura 9 los operadores 'muta uno' y 'muta variable' se aprecian más compactos en sus valores promedios. 'segmentos aleatorios' y 'mutación clásica' exhiben mayor cantidad promedio de individuos diferentes. Estos resultados son parecidos en los casos 2b y 3b. Con el operador de mutación aislado los individuos tienden a ser más parecidos en su genotipo; en este caso, una tasa de cruce alta, introduce mayor variabilidad que la proporción de ruido.

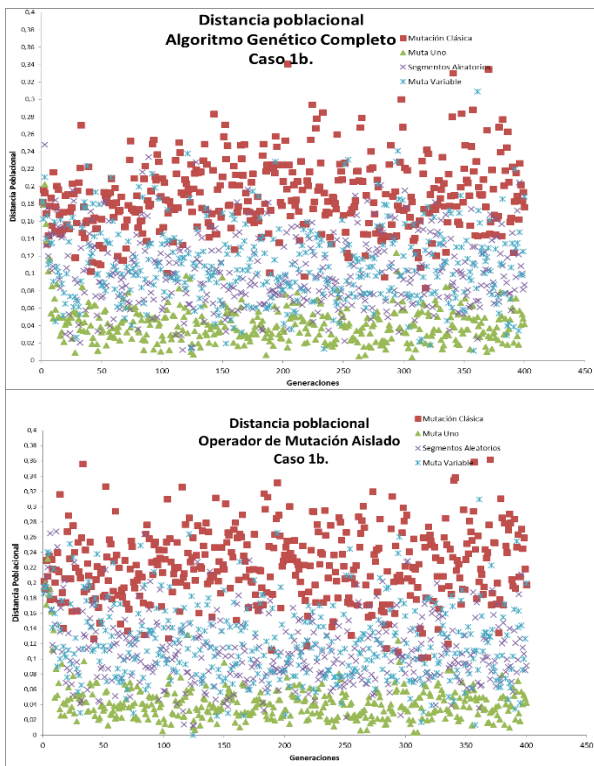


Figura 6: Distancia Poblacional Promedio - Caso 1b

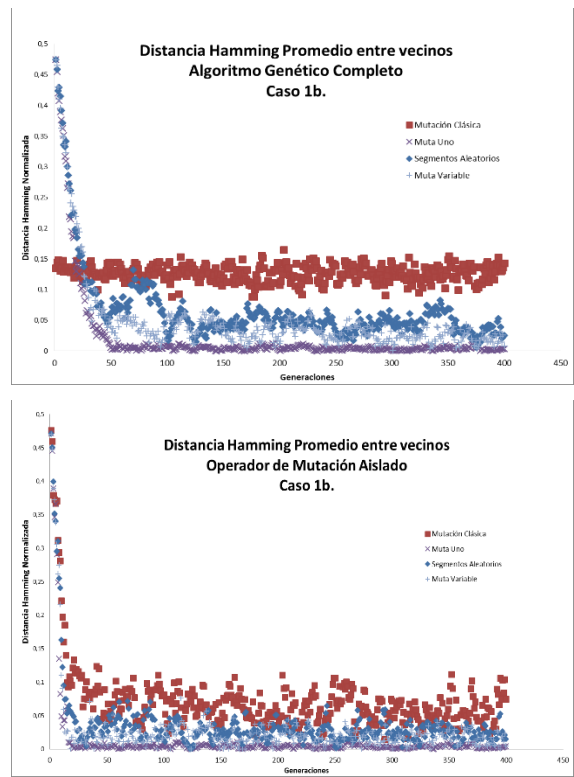


Figura 8: Distancia Hamming Promedio entre Vecinos - Caso 1b

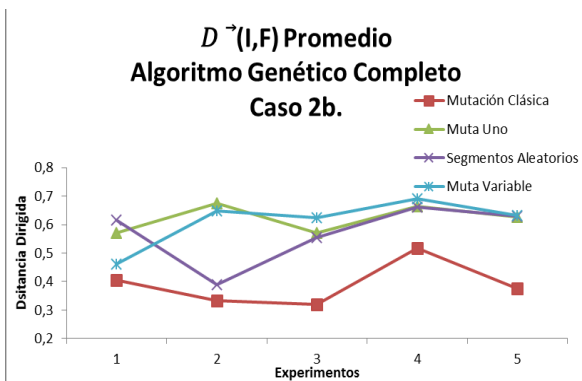
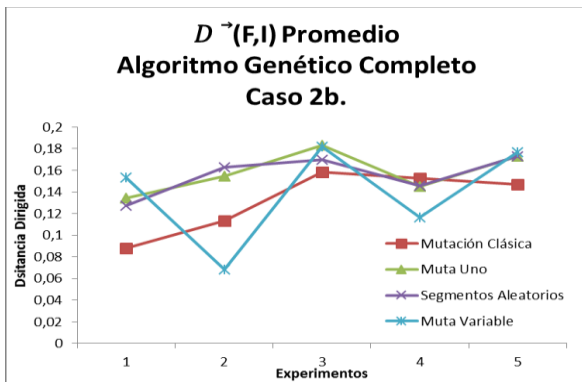


Figura 7: Distancia Dirigida Promedio entre las Poblaciones I y F. Algoritmo Genético Completo - Caso 2b

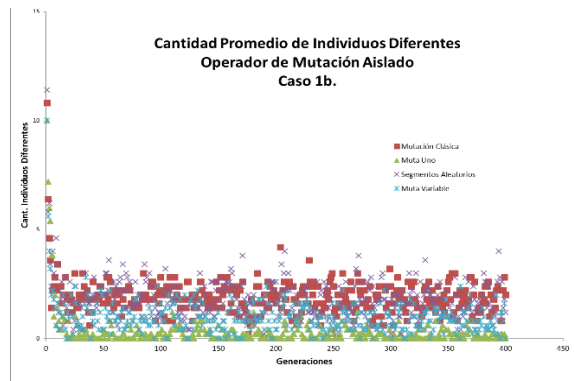
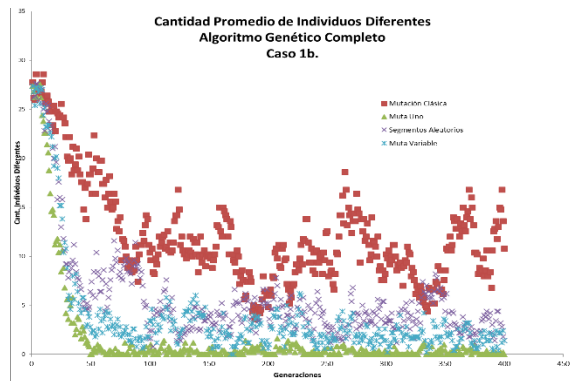


Figura 9: Cantidad Promedio de Individuos Diferentes - Caso 1b

C. Caso c

Se probó $f_1(x, y)$ con $(x, y) \in [-2, 6]$, $f_2(x, y)$ con $(x, y) \in [-6, 8]$, $p_{cruce} = 0.8$, $p_{mutac} = 0.3$, $ruidomutac = 0.2$. Los expertos recomiendan utilizar una tasa de mutación muy pequeña menor a 0.05, por tanto se considera 0.3 una alta tasa de mutación.

En la Figura 10 se observa que los operadores 'segmentos aleatorios' y 'muta uno' obtuvieron una tendencia hacia el óptimo pero con fluctuaciones. El operador de peor comportamiento fue 'mutación clásica' ya que hay aproximadamente un 30% de individuos elegidos que un 20% de probabilidad de que cada alelo mute, lo cual produce mayor cantidad de ruido poblacional que en consecuencia, lo aleja del promedio hacia el óptimo. En este caso tampoco ayuda a la convergencia, la variabilidad adicional introducida por el operador de cruce; por ello no se reportan los resultados del AGS completo. En términos generales tienen tendencia hacia el óptimo pero no se alcanza la convergencia deseada.

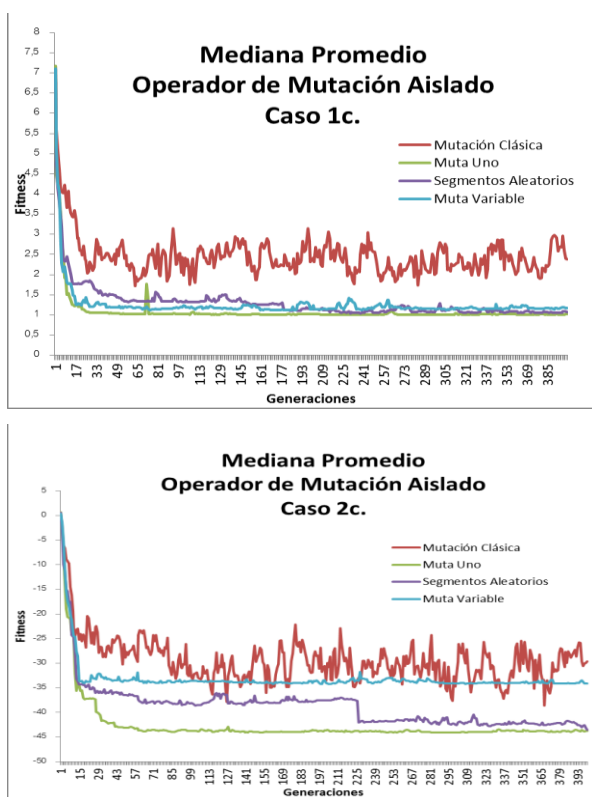


Figura 10: Mediana Promedio - Caso c

En la Figura 11 la distancia poblacional tiene un comportamiento más compacto que el mostrado en el caso b, y el operador 'mutación clásica' imprime mayor distancia entre poblaciones. La distancia dirigida entre las dos poblaciones se mantuvo relativamente similar en comparación con el caso b (Figura 12).

La distancia Hamming promedio entre vecinos de la Figura 13 muestra menos fluctuaciones que en el caso 1b y con mayor parecido al caso 1a por el ruido introducido en los cromosomas mutados. Pese a que tiene alta tasa de mutación, la fluctuaciones son más suaves tendientes a la disminución. Esto indica que la variedad genética se puede acotar, sin

importar la función objetivo, manteniendo constantes las demás condiciones.

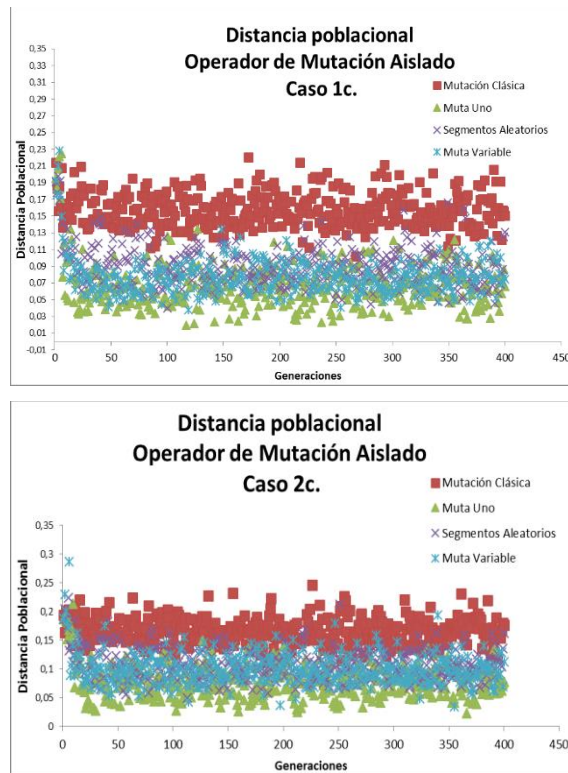


Figura 11: Distancia Poblacional Promedio - Caso c

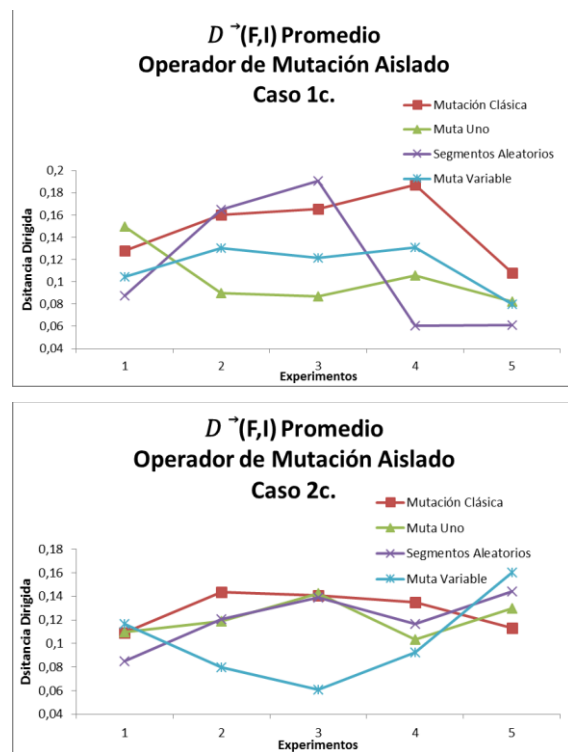


Figura 12: Distancia Dirigida Promedio Población F a I - Caso c

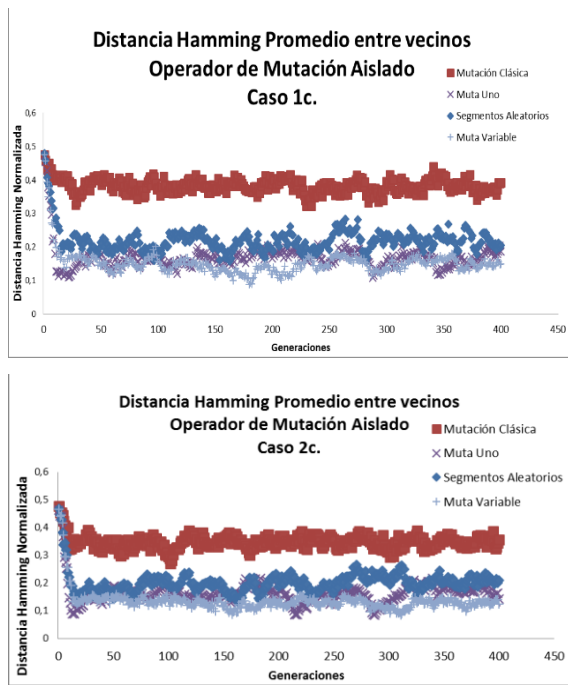


Figura 13: Distancia Hamming Promedio entre vecinos - Caso c

Comparando con las gráficas del caso b, los operadores de la Figura 14 contienen individuos más separados oscilando entre 5 y 15, observándose incluso más dispersión en el AGS completo. En promedio la ‘mutación clásica’ sigue exhibiendo mayor cantidad de individuos diferentes.

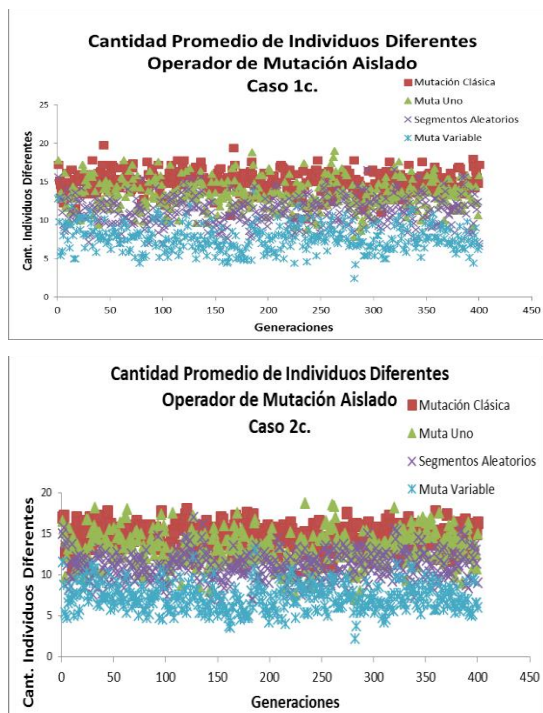


Figura 14: Cantidad Promedio de Individuos Diferentes - Caso c

D. Caso d

Se probó f_1 con $(x,y) \in [-2,6]$, $p_{cruce}=0.8$, $p_{mutac}=0.5$, $ruidomutac=0.7$ y f_3 con $(x,y) \in [-6,6]$, $p_{mutac}=0.3$, $ruidomutac=0.7$.

La distancia Hamming entre individuos es más alta en promedio que en los casos anteriores (entre 0.2 y 0.5), y lo mismo sucede en el cálculo de la cantidad de individuos diferentes: entre 5 y 30 individuos diferentes para las funciones estudiadas. La distancia poblacional se mantuvo en un rango constante por el hecho de que los alelos continuamente cambian de valor y hace que los individuos se mantengan a una misma distancia entre sí (Figura 15).

Para el cálculo de la convergencia, los operadores que lograron acercarse al mínimo óptimo fueron los que introducían menos variación: ‘muta uno’ y ‘muta variable’, pero en términos generales el algoritmo no convergió demostrando que altas tasas de mutación combinadas con una alta proporción de ruido convierten al algoritmo en una búsqueda prácticamente aleatoria.

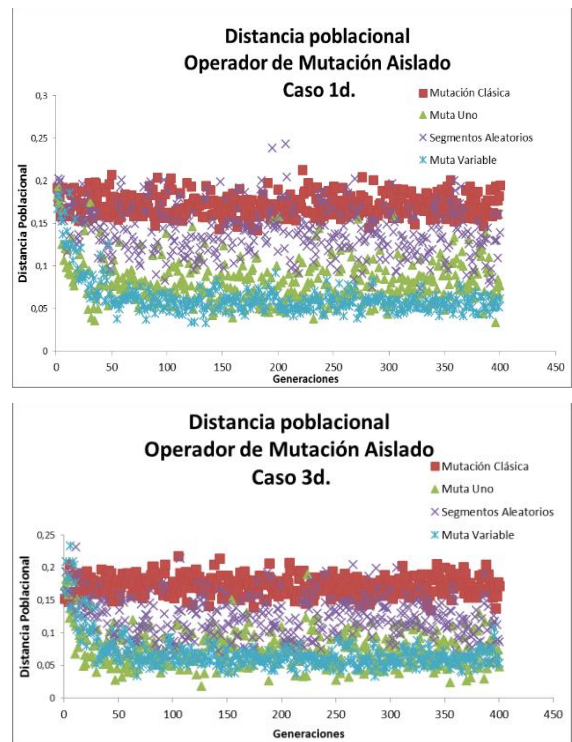


Figura 15: Distancia Poblacional Promedio - Caso d

VI. CONCLUSIONES

Se compararon los resultados de diferentes operadores de mutación mediante la aplicación de medidas que estudian la variedad genética y la convergencia de poblaciones compuestas por alelos binarios. Se diseñó una novedosa pseudo-métrica de distancia poblacional entre Generaciones, cuyo método es flexible y dinámico, para evaluar empíricamente la evolución temporal de la variedad entre poblaciones, lo cual podría extenderse a cromosomas no binarios.

En el estudio de variabilidad genética, los resultados indican que no hay una diferencia significativa entre tener o no el operador de cruce para tasas de mutación y ruido bajas. En el caso a, una alta proporción de cruce beneficia la variabilidad por encima de la mutación ya que ésta última se presenta en muy pocos individuos de la población y muta pocos alelos del cromosoma. En todas las funciones objetivo se logró la convergencia.

Con altas tasas de mutación y ruido (caso d) el algoritmo se convierte en una búsqueda prácticamente aleatoria y difícilmente se alcanza el criterio de convergencia. Queda al margen lo que puedan hacer el operador de selección y cruce.

Por otro lado, los resultados presentados por el operador de mutación aislado del proceso evolutivo con una tasa medianamente alta, el comportamiento de cada operador, a pesar de no ser el ideal, se mantiene alrededor de un mínimo local cercano al óptimo. Esto sugiere, coincidiendo con De Jong [6], un algoritmo genético que al principio sólo contemple mutación y al obtener un espacio solución candidato, incorpore el cruce y disminuya la mutación, para lograr una convergencia más rápida y dar soluciones a diferentes problemas de optimización considerados intratables. Por otro lado, manteniendo el operador de mutación aislado, se observó menor cantidad de individuos diferentes concentrados en un rango menor.

En cuanto a la pseudo-métrica de distancia poblacional dirigida, los resultados muestran que la población final 'F' se parece más a la inicial 'I' que al contrario. Se puede interpretar como una consecuencia del proceso evolutivo al que es sometida la población. En el cálculo de la distancia dirigida entre las poblaciones 'I' y 'F', se observó que a mayor proporción de ruido menores fluctuaciones se presentan en cada uno de los operadores.

Para el caso de altas tasas de mutación y ruido la distancia poblacional global es más compacta, a diferencia de bajas tasas de mutación donde la distancia poblacional es mucho más dispersa. Las sucesivas mutaciones hacen que la población vuelva a una configuración anterior que mantiene una distancia global y entre individuos relativamente estable. Con estos resultados, es posible establecer una cota de variabilidad genética, para cualquier función objetivo, manteniendo constantes ciertas condiciones.

A continuación se plantean conclusiones para cada uno de los operadores.

Mutación Clásica: A pesar de introducir cambios importantes en el cromosoma (reflejado en el fitness promedio) la mediana refleja cierta tendencia de la población hacia un mínimo, lo que hace que, por un lado converja, y por otro que explore nuevas posibilidades en busca del óptimo. Con una tasa de mutación medianamente alta y ruido conservador el operador mutación clásica por sí solo, exhibe buenos resultados en el proceso evolutivo. Posteriormente, hace falta el operador cruce para acercar la solución al óptimo.

En general se ve afectado por la cantidad de ruido en los cromosomas mutados y por ello no logra en varios casos alcanzar la convergencia utilizando la mediana mínima por experimento. Los resultados indican que los individuos se mantienen con una variabilidad relativamente constante y menor en comparación a los demás operadores.

Muta Uno: Tiene un buen desempeño en términos generales y en casi todos los casos alcanza la convergencia. Presenta el peligro de caer en un mínimo local por ser el operador de

menor ruido en los cromosomas mutados, como lo reflejan los resultados de Mediana Mínima por caso. En cuanto a la convergencia, tiene valores de mediana y media bastante cercanos al óptimo y en cuanto a variedad genética, en términos generales, es el operador de menor distancia poblacional global, distancia Hamming entre vecinos, cantidad promedio de individuos diferentes. Cuando existe una tasa alta de mutación, es similar al operador 'muta variable' en las medidas de variedad genética.

Muta Variable: Presenta un pobre desempeño por afectar solo una dimensión del espacio solución; esto hace cambiar la búsqueda en muchas direcciones sin lograr la convergencia.

Segmentos Aleatorios: Es una mutación intermedia entre muta uno y mutación clásica, lo cual produce buenos resultados en la convergencia y no tiene alto impacto en los individuos de la población cuando se aplica cierta cantidad de ruido. Mantiene un espaciamiento relativamente constante en los genotipos de los individuos. Tiene comportamiento similar a 'mutación clásica' en el cálculo de la media y la mediana. En cuanto a la variedad genética y distancia poblacional, con tasas de mutación conservadoras, exhibe un comportamiento similar a 'muta variable'. Para alta cantidad de ruido, el comportamiento en cuanto a variedad genética es similar a 'mutación clásica'. Los resultados de Mediana Mínima por experimento reflejan que para una tasa baja, el operador es robusto al alcanzar la convergencia.

En trabajos futuros se pretende utilizar análisis espectral que arroje indicadores que permitan medir la variedad poblacional y comparar estos resultados con operadores de mutación adaptativos. Además extender las métricas a poblaciones no binarias.

REFERENCIAS

- [1] J. H. Holland, *Outline for a Logical theory of Adaptive Systems*. Journal of the ACM (JACM), vol. 9, no. 3, pp. 297-314, 1962.
- [2] T. Bäck and M. Schütz, *Intelligent Mutation Rate Control in Canonical Genetic Algorithms* in Foundations of Intelligent Systems, Springer, pp. 158-167, 1996.
- [3] T. C. Fogarty, *Varying the Probability of Mutation in the Genetic Algorithm* in Proceedings of the third International Conference on Genetic Algorithms, pp. 104-109, 1989.
- [4] K. A. De Jong, *Analysis of the Behavior of a Class of Genetic Adaptive Systems*. 1975.
- [5] F. Gray, *Pulse Code Communication*. Mar, 1953.
- [6] D. E. Golberg, *Genetic algorithms in Search, Optimization, and Machine Learning*. Addison Wesley, vol. 1989.
- [7] T. C. Fogarty, *Varying the Probability of Mutation in the Genetic Algorithm* in Proceedings of the Third International Conference on Genetic Algorithms, pp. 104-109, 1989.
- [8] T. BVack, *Optimal Mutation Rates in Genetic Search* in Proceedings of the Fifth International Conference on Genetic Algorithms, pp. 2-8, 1993.
- [9] G. Ochoa, I. Harvey, and H. Buxton, *Optimal Mutation Rates and Selection Pressure in Genetic Algorithms* in GECCO, pp. 315-322, 2000.
- [10] A. Papadopoulos, *Metric Spaces, Convexity and Nonpositive Curvature*. European Mathematical Society, vol. 6, 2005.