Modelado Matemático del Algoritmo UMDA$_C$ con Selección por Torneo Aplicado a Funciones Lineales

C. González, J. A. Lozano, P. Larrañaga

Resumen— **Este artículo está dedicado al estudio teórico del comportamiento del algoritmo UMDA$_c$ (Univariate Marginal Distribution Algorithm for continuous domains) en espacios de dimensión n. Para ello estudiaremos la aplicación de dicho algoritmo, con selección por torneo, a la minimización de la función lineal $G(\mathbf{x}) = a_0 + \sum_{i=1}^{n} a_i x_i$, con $\mathbf{x} = (x_1, \ldots, x_n) \in I\!R^n$ y $a_i \in I\!R$, $i = 0, 1, \ldots, n$.**

En el presente trabajo conseguimos, por un lado, modelar matemáticamente el algoritmo en cuestión cuando se hacen un número infinito de torneos a cada paso, y por otro, utilizar dicho modelado para demostrar el mal comportamiento que presenta en la función lineal $G_1(\mathbf{x}) = \sum_{i=1}^{n} x_i$, con $\mathbf{x} = (x_1, \ldots, x_n) \in I\!R^n$.

Palabras clave— **UMDA$_c$, funciones lineales, selección por torneo, modelado matemático, EDAs.**

I. INTRODUCCIÓN

Los algoritmos de estimación de distribuciones (Estimation of Distribution Algorithms, EDAs) [1], [2] constituyen una nueva y prometedora área dentro de la Computación Evolutiva. Esta nueva clase de algoritmos surge como una generalización de los Algoritmos Genéticos (AGs), con el ánimo de mejorar el mal comportamiento que éstos últimos presentan en ciertos problemas deceptivos.

Introducidos por Mühlenbein y Paaβ [2], los EDAs son un ejemplo de heurísticos estocásticos basados en poblaciones de individuos, donde cada individuo codifica una posible solución del problema de optimización. Dichas poblaciones evolucionan en generaciones sucesivas según avanza la búsqueda, al igual que en los AGs. La diferencia fundamental entre los AGs y los EDAs reside en el reemplazamiento de los operadores de cruce y mutación (esenciales en los AGs a la hora de generar nuevas poblaciones), por el aprendizaje y muestreo de la distribución de probabilidad que siguen los mejores individuos de la población. De esta manera los EDAs capturan y explotan de manera eficiente las relaciones entre las variables del problema.

La dificultad fundamental que surge con ésta aproximación radica en la estimación de la distribución de probabilidad conjunta asociada a la base de datos que contiene a los individuos seleccionados. Para superar éste problema, algunos autores han propuesto diferentes algoritmos donde se simplifican las asunciones referentes a las dependencias entre las variables de la distribución de probabilidad conjunta. En los siguientes artículos se pueden encontrar revisiones de diferentes trabajos en los campos combinatorio y numérico [3], [4], [5], [6].

Hasta el momento la mayoría de trabajos que existen en la literatura referentes a los EDAs, están dedicados a aplicaciones o son nuevas variantes . Se pueden encontrar muy pocos trabajos dedicados a modelarlos matemáticamente.

Comenzamos por citar trabajos en el campo discreto. Los más generales que se pueden encontrar son dos, por un lado, el trabajo de González y col. [7], en el que, además de ofrecerse una visión unificada de los aspectos teóricos de los EDA discretos existentes en la literatura, se presenta un teorema general de convergencia para los mismos. Por el otro, Mühlenbein y col. [8] introducen un EDA que usa selección de Boltzmann: Boltzmann Estimation of Distribution Algorithm (BEDA). En dicho trabajo muestran la convergencia de un BEDA general para poblaciones infinitas. Tambien en el campo discreto, pero ahora ya citando trabajos que se dedican al análisis de un algoritmo en concreto, tenemos: respecto al estudio del UMDA (Univariate Marginal Distribution Algorithm), el trabajo de Mühlenbein y Mahnig [9], donde se prueba que el UMDA con población infinita y selección proporcional termina en óptimos locales. Podemos también citar varios artículos que analizan otro algoritmo EDA discreto, el PBIL (Population Based Incremental Learning Algorithm). Höhfeld y Rudolph [10] estudian el comportamiento del PBIL en funciones lineales. González y col. [11], modelan el algoritmo mediante cadenas de Markov para demostrar que la convergencia de PBIL, aplicado a la función $OneMax$ en dimensión dos, tiene una fuerte dependencia de los parámetros iniciales, mientras que en González y col. [12], asocian un sistema dinámico al PBIL, demostrando que el algoritmo sigue las iteraciones de dicho sistema dinámico y llegando a la conclusión de que los puntos del espacio de búsqueda son puntos fijos del sistema dinámico mientras que los óptimos locales son los puntos fijos estables. En Berny [13], se demuestra que PBIL puede derivarse de un sistema dinámico gradiente, además se realiza un análisis de la estabilidad de dicho sistema, concluyendo que PBIL únicamente puede converger a soluciones locales.

Pasando al campo continuo tan sólo podemos encontrar un trabajo de corte teórico Berny [14], en el que el autor analiza el PBIL$_c$ (Population Based

Intelligent Systems Group, Departamento de Ciencias de la Computación e Inteligencia Artificial, Facultad de Informática, M. de Lardizabal Pasalekua 1, 20009 - San Sebastián (España) {ccpgomoc,ccploalj,ccplamup}@si.ehu.es.

Incremental Learning Algorithm for continuous domains), aportando un estudio análogo al realizado en el caso discreto, pero ahora para funciones reales continuas, y sin dar resultados sobre estabilidad. Nosotros, en este artículo, intentamos engrosar la lista anterior analizando desde un punto de vista teórico el algoritmo UMDA$_c$.

El UMDA constituye la versión más simple de los EDAs. Fué introducido en su vertiente discreta por Mühlenbein [15]. La primera versión continua del algoritmo fué dada por Larrañaga y col. [4], [6]. Éste algoritmo supone independencia entre las variables, por lo tanto se asume que la distribución de probabilidad n-dimensional factoriza como el producto de n distribuciones de probabilidad univariadas e independientes.

Es muy usual suponer que la función de densidad conjunta es una distribución normal n-dimensional, que factoriza como el producto de n normales unidimensionales e independientes. De hecho, en éste trabajo haremos ésta suposición.

Con éste artículo pretendemos conocer el comportamiento del algoritmo UMDA$_c$ con selección por torneo y para ello hemos modelado matemáticamente la aplicación del mismo a la minimización de funciones lineales en dimensión n. La utilización de éstas funciones da lugar al modelado del algoritmo cuando nos encontramos lejos del óptimo.

El trabajo está estructurado del siguiente modo. En la Sección 2 se da una visión más detallada del UMDA$_c$ con selección por torneo. La Sección 3 está dedicada al modelado matemático del algoritmo. La Sección 4 incluye determinados cálculos que son de gran importancia a la hora del estudio teórico del mismo, resumiéndose los resultados en la Sección 5. Finalmente, las conclusiones, así como el trabajo futuro, se exponen en la Sección 6.

II. El algoritmo UMDA$_c$ con selección por torneo

En ésta sección detallaremos el funcionamiento del algoritmo UMDA$_c$ con selección por torneo.

El algoritmo funciona como sigue. A cada paso t se mantiene una variable aleatoria n-dimensional $\mathbf{X}^t = (X_1^t, \ldots, X_n^t)$ en la que cada una de sus componentes está distribuida como una normal unidimensional:

$$\left(f_{\mathcal{N}(\mu_1^t,\sigma_1^t)}, f_{\mathcal{N}(\mu_2^t,\sigma_2^t)}, \ldots, f_{\mathcal{N}(\mu_n^t,\sigma_n^t)}\right), \quad (1)$$

dónde $f_{\mathcal{N}(\mu_i^t,\sigma_i^t)}(x_i) = \frac{1}{\sqrt{2\pi}\sigma_i^t} e^{-(x_i-\mu_i^t)^2/2(\sigma_i^t)^2}$, $i = 1, \ldots, n$, dicho de otro modo $f_{\mathcal{N}(\mu_i^t,\sigma_i^t)}(x_i)$ está denotando la función de densidad de una normal unidimensional de media μ_i^t y desviación típica $(\sigma_i^t)^2$.

Se muestrea la variable aleatoria antes descrita, obteniéndose dos individuos. Se lleva a cabo una selección por torneo, es decir, el mejor de ellos pasa a ser seleccionado. Repitiendo este proceso N veces se obtiene la población de los seleccionados, y a partir de ella las medias y desvíaciones típicas para conseguir nuevas distribuciones unidimensionales en el paso $t+1$. Un pseudo-código del algoritmo descrito anteriormente puede verse en la Figura 1.

Obtener aleatoriamente una distribucion de probabilidad normal para cada una de las variables.
mientras no convergencia **hacer**
  begin
  for$(j = 1; j \leq N; j{+}{+})$
    begin
      Muestrando $\mathbf{X}^t$ obtener
      2 individuos :
        $\mathbf{x}_{1,j}^t = (x_{1,j}^{1,t}, \ldots, x_{1,j}^{n,t})$
        $\mathbf{x}_{2,j}^t = (x_{2,j}^{1,t}, \ldots, x_{2,j}^{n,t})$
      Evaluar $\mathbf{x}_{1,j}^t$, $\mathbf{x}_{2,j}^t$
      Selecionar el mejor de ellos:
        $\mathbf{x}_{(1:2),j}^t =$
        $= argmin_{\mathbf{x}}\{G(\mathbf{x}_{1,j}^t), G(\mathbf{x}_{2,j}^t)\}$
    end
  for$(i = 1; i \leq n; i{+}{+})$
    begin
      Estimar los parámetros de las nuevas funciones de densidad:
        $\mu_i^{t+1} = \frac{\sum_{j=1}^N x_{(1:2),j}^{i,t}}{N}$
        $\sigma_i^{t+1} = \sqrt{\frac{\sum_{j=1}^N (x_{(1:2),j}^{i,t} - \mu_i^{t+1})^2}{N}}$
    end
  end

Fig. 1. Pseudo-código UMDA$_c$ selección torneo.

Nuestro objetivo es conocer cómo cambian las distribuciones de probabilidad de un paso a otro, para saber cómo evolucionan μ_i^t y σ_i^t cuando t crece.

III. Modelado matemático del algoritmo

Se trata de modelar el algoritmo UMDA$_c$ con selección por torneo para problemas de optimización continuos de n variables. Modelaremos el caso en que, a cada paso se realizan infinitos torneos. La función a minimizar es:

$$\begin{array}{rccl} G: & I\!R^n & \longrightarrow & I\!R \\ & \mathbf{x} & \longmapsto & a_0 + \sum_{i=1}^n a_i x_i \ . \end{array}$$

Suponemos, tal y como se ha dicho antes, que a cada paso t cada variable X_i^t, $i = 1, \ldots, n$ está distribuida como una normal unidimensional, es decir cada variable X_i^t tiene asociada la siguiente función de densidad:

$$f_{\mathbf{X}_i^t}(x_i) = f_{\mathcal{N}(\mu_i^t,\sigma_i^t)}(x_i) = \frac{1}{\sqrt{2\pi}\sigma_i^t} e^{-(x_i-\mu_i^t)^2/2(\sigma_i^t)^2}$$

donde $f_{\mathbf{X}_i^t}(x_i)$ denota la función de densidad de la variable aleatoria X_i^t. Dado que estamos trabajando con UMDA$_c$ dichas variables son independientes entre sí. De esta manera a cada paso t tendremos una variable aleatoria n-dimensional $\mathbf{X}^t = (X_1^t, \ldots, X_n^t)$ que sigue una distribución de la forma:

$$\left(f_{\mathcal{N}(\mu_1^t,\sigma_1^t)}, f_{\mathcal{N}(\mu_2^t,\sigma_2^t)}, \dots, f_{\mathcal{N}(\mu_n^t,\sigma_n^t)}\right). \quad (4)$$

Durante el razonamiento y para aligerar la notación, cada una de las funciones de densidad asociadas a cada variable X_i^t será denotada por:

$$f_i(x_i) = f_{X_i^t}(x_i) = f_{\mathcal{N}(\mu_i^t,\sigma_i^t)}(x_i), \text{ con } i \in \{1, 2 \dots, n\}. \quad (5)$$

De la misma manera cada función de distribución asociada se escribirá:

$$F_i(x_i) = \int_{-\infty}^{x_i} f_i(s)ds, \text{ con } i \in \{1, 2 \dots, n\}. \quad (6)$$

Usaremos la notación habitual tanto en el caso de la función de densidad de la normal estandar:

$$\phi(x) = f_{\mathcal{N}(\mu=0,\sigma=1)}(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}, \quad (7)$$

como en el caso de su función de distribución asociada:

$$\Phi(x) = \int_{-\infty}^{x} \phi(s)ds. \quad (8)$$

A cada paso t se considera la variable aleatoria $\mathbf{X}_{(1:2)}^t$, es decir la variable aleatoria del mejor de 2 variables del tipo $\mathbf{X}^t$. Se calculan:

$$\begin{aligned} \mathsf{E}[\mathbf{X}_{(1:2)}^t] &= \left(\mathsf{E}[X_{(1:2),1}^t], \dots, \mathsf{E}[X_{(1:2),n}^t]\right) \\ \mathsf{Var}[\mathbf{X}_{(1:2)}^t] &= \left(\mathsf{Var}[X_{(1:2),1}^t], \dots, \mathsf{Var}[X_{(1:2),n}^t]\right) , \end{aligned}$$

y una vez hecho esto se calculan las nuevas distribuciones en el paso $t+1$, de manera que cada X_i^{t+1} estará distribuida como una normal del tipo:

$$\mathcal{N}^{t+1}\left(\mu_i^{t+1} = \mathsf{E}[X_{(1:2),i}^t], \sigma_i^{t+1} = \sqrt{\mathsf{Var}[X_{(1:2),i}^t]}\right) \quad \text{con} \quad i = 1, \dots, n \ . \quad (9)$$

Lógicamente y dado que queremos modelar el comportamiento de éste algoritmo cuando t crece, el siguiente paso a dar consiste en conocer explícitamente las expresiones de μ_i^{t+1} y σ_i^{t+1} dados μ_i^t y σ_i^t, con $i = 1, \dots, n$ para, posteriormente, analizar las sucesiones $\{\mu_i^t\}_t$ y $\{\sigma_i^t\}_t$ con $t \in I\!N$ y ver cómo evolucionan cuando aumenta el número de iteraciones, es decir estudiar los límites:

$$\lim_{t\to\infty} \mu_i^t \quad (10)$$

$$\lim_{t\to\infty} \sigma_i^t \ . \quad (11)$$

El cálculo de μ_i^{t+1} y σ_i^{t+1} exige primero encontrar la función de densidad asociada al mejor individuo de cada torneo. Dicha función de densidad será denotada por $f_{(1:2)}^t(x_1, \dots, x_n)$ y dependerá de G, la función objetivo que estamos considerando.

A. *Cálculo de $f_{(1:2)}^t(x_1, \dots, x_n)$*

Haremos el cálculo de la función de densidad $f_{(1:2)}^t(x_1, \dots, x_n)$ de la forma clásica. Primero obtendremos la función de distribución asociada, que denotaremos por $F_{(1:2)}^t(x_1, \dots, x_n)$, para más tarde, derivarla y conseguir la función de densidad $f_{(1:2)}^t(x_1, \dots, x_n)$.

Sea $\mathbf{X}_1^t = (X_{1,1}^t, \dots, X_{1,n}^t)$ la variable aleatoria correspondiente al primer individuo que se obtiene para el torneo en el paso t y $\mathbf{X}_2^t = (X_{2,1}^t, \dots, X_{2,n}^t)$ la correspondiente al segundo individuo. Entonces, $F_{(1:2)}^t(x_1, \dots, x_n)$ será:

$$\begin{aligned} &F_{(1:2)}^t(x_1, \dots, x_n) \\ &= P\left((X_{(1:2),1}^t, \dots, X_{(1:2),n}^t) \le (x_1, \dots, x_n)\right) \end{aligned} \quad (12)$$

Expresamos la variable aleatoria asociada al mejor individuo de cada torneo $\mathbf{X}_{(1:2)}^t$ como una suma de variables aleatorias para facilitar el cálculo:

$$\mathbf{X}_{(1:2)}^t = \mathbf{X}_1^t \cdot 1_{\{G(\mathbf{X}_1^t) \le G(\mathbf{X}_2^t)\}} + \mathbf{X}_2^t \cdot 1_{\{G(\mathbf{X}_1^t) > G(\mathbf{X}_2^t)\}} , \quad (13)$$

donde la variable aleatoria 1_A es el indicador del suceso A, y por lo tanto:

$$1_A(x) = \begin{cases} 1 & \text{si } x \in A \\ 0 & \text{en otro caso.} \end{cases} \quad (14)$$

Escribimos el suceso del que hay que calcular la probabilidad en (12) como una unión de sucesos:

$$\begin{aligned} &\{(X_{(1:2),1}^t, \dots, X_{(1:2),n}^t) \le (x_1, \dots, x_n)\} \\ &= \{U_1^t \cap U_2^t\} \cup \{V_1^t \cap V_2^t\} \ , \end{aligned} \quad (15)$$

donde:

$$\begin{aligned} U_1^t &= \{(X_{1,1}^t, \dots, X_{1,n}^t) \le (x_1, \dots, x_n)\} \\ U_2^t &= \{G(X_{1,1}^t, \dots, X_{1,n}^t) \le G(X_{2,1}^t, \dots, X_{2,n}^t)\} \\ V_1^t &= \{(X_{2,1}^t, \dots, X_{2,n}^t) \le (x_1, \dots, x_n)\} \\ V_2^t &= \{G(X_{1,1}^t, \dots, X_{1,n}^t) > G(X_{2,1}^t, \dots, X_{2,n}^t)\}. \end{aligned}$$

Denotamos U^t al suceso $U_1^t \cap U_2^t$ y V^t al suceso $V_1^t \cap V_2^t$, y dado que:

$$\{(X_{(1:2),1}^t, \dots, X_{(1:2),n}^t) \le (x_1, \dots, x_n)\} = \{U^t\} \cup \{V^t\},$$

podemos decir que :

$$F_{(1:2)}^t(x_1, \dots, x_n) = P(U^t) + P(V^t) \ . \quad (16)$$

Teniendo en cuenta que $P(U^t) = P(V^t)$, basta con obtener $P(U^t)$, para ello calcularemos la probabilidad condicionada $P(U^t | X_{1,1}^t = x_{1,1}, \dots, X_{1,n}^t = x_{1,n})$ y luego integraremos en las variables condicionantes:

$$\begin{aligned} &P(U^t | X_{1,1}^t = x_{1,1}, \dots, X_{1,n}^t = x_{1,n}) \\ &= P\left(G(X_{2,1}^t, \dots, X_{2,n}^t) \ge G(x_{1,1}, \dots, x_{1,n})\right) \ , \\ &\text{si} \quad x_{1,1} \le x_1, \dots, x_{1,n} \le x_n \quad (17) \\ &P(U^t | X_{1,1}^t = x_{1,1}, \dots, X_{1,n}^t = x_{1,n}) = 0 \ , \\ &\text{en otro caso} \quad (18) \end{aligned}$$

Entonces:

$$
\begin{aligned}
P(U^t) &= \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} P(U^t | X_{1,1}^t = x_{1,1}, \ldots, X_{1,n}^t = x_{1,n}) \cdot \\
&\quad f_1^t(x_{1,1}) \cdot \ldots \cdot f_n^t(x_{1,n}) dx_{1,1} \ldots dx_{1,n} \\
&= \int_{-\infty}^{x_1} \cdots \int_{-\infty}^{x_n} P(G(X_{2,1}^t, \ldots, X_{2,n}^t) \geq G(x_{1,1}, \ldots, x_{1,n})) \cdot \\
&\quad \cdot f_1^t(x_{1,1}) \cdot \ldots \cdot f_n^t(x_{1,n}) dx_{1,1} \ldots dx_{1,n} \\
&= P(V^t)
\end{aligned}
$$

Por lo tanto, la función de distribución de $\mathbf{X}_{(1:2)}^t$ será:

$$
\begin{aligned}
& F_{(1:2)}^t(x_1, \ldots, x_n) \\
&= P\left((X_{(1:2),1}^t, \ldots, X_{(1:2),n}^t) \leq (x_1, \ldots, x_n) \right) \\
&= 2 \cdot \\
&\cdot \int_{-\infty}^{x_1} \cdots \int_{-\infty}^{x_n} P(G(X_{2,1}^t, \ldots, X_{2,n}^t) \geq G(x_{1,1}, \ldots, x_{1,n})) \cdot \\
&\cdot f_1^t(x_{1,1}) \cdot \ldots \cdot f_n^t(x_{1,n}) dx_{1,1} \ldots dx_{1,n} ,
\end{aligned}
$$

y derivando la expresión anterior obtenemos la función de densidad:

$$
\begin{aligned}
& f_{(1:2)}^t(x_1, \ldots, x_n) = 2 f_1^t(x_1) \cdot \ldots \cdot f_n^t(x_n) \cdot \\
& \quad \cdot P(G(X_{2,1}^t, \ldots, X_{2,n}^t) \geq G(x_1, \ldots, x_n)) \quad (19)
\end{aligned}
$$

para simplificar la notación escribiremos:

$$
P(G(X_{2,1}^t, \ldots, X_{2,n}^t) \geq G(x_1, \ldots, x_n)) = A^t \left(a_0 + \sum_{i=1}^n a_i x_i \right) \tag{20}
$$

entonces:

$$
f_{(1:2)}^t(x_1, \ldots, x_n) = 2 \prod_{i=1}^n f_i^t(x_i) A^t \left(a_0 + \sum_{i=1}^n a_i x_i \right). \tag{21}
$$

Tal y como hemos comentado anteriormente, si queremos conocer cómo se comporta el algoritmo descrito en la sección anterior es necesario calcular:

$$
\mu_i^{t+1} = \mathsf{E}[X_{(1:2),i}^t] \tag{22}
$$

$$
\sigma_i^{t+1} = \sqrt{\mathsf{Var}[X_{(1:2),i}^t]} , \tag{23}
$$

con $i = 1, \ldots, n$ para poder obtener $\mathcal{N}^{t+1} \left(\mu_i^{t+1} = \mathsf{E}[X_{(1:2),i}^t], \sigma_i^{t+1} = \sqrt{\mathsf{Var}[X_{(1:2),i}^t]} \right)$, $i = 1, \ldots, n$ y poder así comprobar qué ocurre cuando el número de iteraciones crece.

IV. Cálculo de μ_i^{t+1} y σ_i^{t+1}

Para obtener μ_i^{t+1} y σ_i^{t+1} necesitamos conocer cada una de las funciones de densidad marginales, $f_{(1:2),i}^t(x_i)$ con $i = 1, \ldots, n$ puesto que:

$$
\begin{aligned}
& (\mu_1^{t+1}, \ldots, \mu_n^{t+1}) \\
&= \left(\int_{-\infty}^{\infty} x_1 f_{(1:2),1}^t(x_1) dx_1, \ldots, \right. \\
& \quad \left. \ldots, \int_{-\infty}^{\infty} x_n f_{(1:2),n}^t(x_n) dx_n \right) .
\end{aligned}
$$

$$
\begin{aligned}
& (\sigma_1^{t+1}, \ldots, \sigma_n^{t+1}) \\
&= \left(\sqrt{\int_{-\infty}^{\infty} x_1^2 f_{(1:2),1}^t(x_1) dx_1 - (\mu_1^{t+1})^2}, \ldots \right. \\
& \quad \left. \ldots, \sqrt{\int_{-\infty}^{\infty} x_n^2 f_{(1:2),n}^t(x_n) dx_n - (\mu_n^{t+1})^2} \right) . \quad (24)
\end{aligned}
$$

Para no complicar la notación, en los cálculos que siguen, eliminaremos el superíndice correspondiente al paso, aunque en la expresión final de μ_i^{t+1} y σ_i^{t+1} si escribiremos los superíndices.

Escribimos la expresión de las densidades marginales:

$$
\begin{aligned}
f_{(1:2),i}(x_i) &= \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f_{(1:2)}(x_1, \ldots, x_n) \cdot \\
& \cdot dx_1, \ldots, dx_{i-1}, dx_{i+1}, \ldots, dx_n \\
&= \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} 2 \prod_{j=1}^n f_j(x_j) A^t \left(a_0 + \sum_{j=1}^n a_j x_j \right) \cdot \\
& \cdot dx_1, \ldots, dx_{i-1}, dx_{i+1}, \ldots, dx_n \\
&= 2 f_i(x_i) h_i(x_i) \ ,
\end{aligned} \tag{25}
$$

siendo:

$$
\begin{aligned}
h_i(x_i) = & \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \prod_{\substack{j=1 \\ j \neq i}}^n f_j(x_j) A^t \left(a_0 + \sum_{j=1}^n a_j x_j \right) \cdot \\
& dx_1, \ldots, dx_{i-1}, dx_{i+1}, \ldots, dx_n \ .
\end{aligned} \tag{26}
$$

A. Cálculo de μ_i^{t+1}

Hacemos los cálculos sólo para la primera de las componentes ya que para el resto serían análogos:

$$
\mu_1^{t+1} = \int_{-\infty}^{\infty} 2 x_1 f_1(x_1) h(x_1) dx_1. \tag{27}
$$

Primeramente necesitamos conocer el valor de $A^t \left(a_0 + \sum_{j=1}^n a_j x_j \right)$:

$$
A^t \left(a_0 + \sum_{j=1}^n a_j x_j \right) = P \left(a_0 + \sum_{j=1}^n a_j X_{2,j}^t \geq a_0 + \sum_{j=1}^n a_j x_j \right) \tag{28}
$$

dado que cada $X_{2,i}^t$ es una variable aleatoria con función de densidad $f_{\mathcal{N}}(\mu_i, \sigma_i)$, sabemos que la variable aleatoria $T = \sum_{j=1}^n a_j X_{2,j}^t$ tiene función de densidad $\mathcal{N} \left(\sum_{j=1}^n a_j \mu_j, \sqrt{\sum_{j=1}^n a_j^2 \sigma_j^2} \right)$. Entonces:

$$
\begin{aligned}
& A^t \left(a_0 + \sum_{j=1}^n a_j x_j \right) \\
& \quad = P \left(\frac{T - \sum_{j=1}^n a_j \mu_j}{\sqrt{\sum_{j=1}^n a_j^2 \sigma_j^2}} \geq \frac{\sum_{j=1}^n a_j (x_j - \mu_j)}{\sqrt{\sum_{j=1}^n a_j^2 \sigma_j^2}} \right)
\end{aligned}
$$

$$= 1 - \Phi\left(\frac{\sum_{j=1}^{n} a_j(x_j - \mu_j)}{\sqrt{\sum_{j=1}^{n} a_j^2\sigma_j^2}}\right) . \quad (29)$$

Ahora podemos calcular $h_1(x_1)$:

$$h_1(x_1) = \int_{\infty}^{\infty} f_n(x_n)\ldots\int_{\infty}^{\infty} f_k(x_k)\ldots\int_{\infty}^{\infty} f_2(x_2)dx_2\ldots dx_n,$$

llamando:

$$g_k(x_1, x_{k+1}, x_{k+2}, \ldots, x_n) = \int_{\infty}^{\infty} f_k(x_k)\ldots\int_{\infty}^{\infty} f_2(x_2)dx_2\ldots dx_k \quad (30)$$

con $k = 2, \ldots, n$, sabemos que:

$$h_1(x_1) = g_n(x_1) \ . \quad (31)$$

Vamos a demostrar por inducción sobre k que:

$$g_k(x_1, x_{k+1}, x_{k+2}, \ldots, x_n) = 1 - \Phi\left(\frac{\sum_{\substack{i=1\\ i\neq 2,\ldots,k}}^{n} a_i(x_i - \mu_i)}{\sqrt{\sum_{i=2}^{k} a_i^2\sigma_i^2 + \sum_{i=1}^{n} a_i^2\sigma_i^2}}\right). \quad (32)$$

Para ello lo demostramos primero para $k = 2$, después tomamos como hipótesis de inducción el caso k y demostramos la validez de (32) para el caso $k+1$.

Antes de demostrar la fórmula para el caso $k = 2$ es necesario introducir tres fórmulas [16] que nos serán de utilidad a la hora de hacer los cálculos:

$$I_0(a, b) = \int_{-\infty}^{\infty} e^{-s^2/2}\Phi(as + b)ds = \sqrt{2\pi}\Phi\left(\frac{b}{\sqrt{1 + a^2}}\right) \quad (33)$$

$$I_1(a, b) = \int_{-\infty}^{\infty} se^{-s^2/2}\Phi(as + b)ds = \frac{a}{\sqrt{1 + a^2}}\exp\left(\frac{-1}{2}\frac{b^2}{1 + a^2}\right) \quad (34)$$

$$I_2(a, b) = \int_{-\infty}^{\infty} s^2e^{-s^2/2}\Phi(as + b)ds = \sqrt{2\pi}\Phi\left(\frac{b}{\sqrt{1 + a^2}}\right) - \frac{a^2b}{(1 + a^2)\sqrt{1 + a^2}}\exp\left(\frac{-1}{2}\frac{b^2}{1 + a^2}\right). \quad (35)$$

Ahora ya podemos verificar que se cumple (32) para $k = 2$:

$$g_2(x_1, x_3, x_4, \ldots, x_n) = \int_{-\infty}^{\infty}\frac{1}{\sqrt{2\pi}\sigma_2}e^{-(x_2-\mu_2)^2/2\sigma_2^2} \cdot \left(1 - \Phi\left(\frac{\sum_{i=1}^{n} a_i(x_i - \mu_i)}{\sqrt{\sum_{i=1}^{n} a_i^2\sigma_i^2}}\right)\right)dx_2 .$$

Efectuando el cambio de variable:

$$\left|\begin{array}{c} \frac{x_2-\mu_2}{\sigma_2} = s \\ dx_2\sigma_2 ds \end{array}\right| ,$$

obtenemos que:

$$g_2(x_1, x_3, x_4, \ldots, x_n) = 1 - \int_{-\infty}^{\infty}\frac{1}{\sqrt{2\pi}}e^{-s^2/2} \cdot \Phi\left(\frac{a_2\sigma_2 s + a_2\mu_2 + \sum_{\substack{i=1\\ i\neq 2}}^{n} a_ix_i - \sum_{i=1}^{n} a_i\mu_i}{\sqrt{\sum_{i=1}^{n} a_i^2\sigma_i^2}}\right)ds$$

$$= 1 - \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{-s^2/2} \cdot \Phi\left(\frac{a_2\sigma_2 s}{\sqrt{\sum_{i=1}^{n} a_i^2\sigma_i^2}} + \frac{\sum_{\substack{i=1\\ i\neq 2}}^{n} a_i(x_i - \mu_i)}{\sqrt{\sum_{i=1}^{n} a_i^2\sigma_i^2}}\right)ds \ .$$

Utilizando el resultado (33):

$$g_2(x_1, x_3, x_4, \ldots, x_n) = 1 - \frac{1}{\sqrt{2\pi}} \cdot I_0\left(\frac{a_2\sigma_2}{\sqrt{\sum_{i=1}^{n} a_i^2\sigma_i^2}}, \frac{\sum_{\substack{i=1\\ i\neq 2}}^{n} a_i(x_i - \mu_i)}{\sqrt{\sum_{i=1}^{n} a_i^2\sigma_i^2}}\right)$$

$$= 1 - \Phi\left(\frac{\sum_{\substack{i=1\\ i\neq 2}}^{n} a_i(x_i - \mu_i)}{\sqrt{a_2^2\sigma_2^2 + \sum_{i=1}^{n} a_i^2\sigma_i^2}}\right) . \quad (36)$$

Por la hipótesis de inducción suponemos que (32) es cierto para k, veamos que también es cierto para $k+1$:

$$g_{k+1}(x_1, x_{k+2}, x_{k+3}, \ldots, x_n) = \int_{-\infty}^{\infty} f_{k+1}(x_{k+1})g_k(x_1, x_{k+1}, x_{k+1}, \ldots, x_n)dx_{k+1}$$

$$= 1 - \int_{-\infty}^{\infty} f_{k+1}(x_{k+1}) \cdot \Phi\left(\frac{\sum_{\substack{i=1\\ i\neq 2,\ldots,k}}^{n} a_i(x_i - \mu_i)}{\sqrt{\sum_{i=2}^{k} a_i^2\sigma_i^2 + \sum_{i=1}^{n} a_i^2\sigma_i^2}}\right)dx_{k+1} \ .$$

Haciendo el siguiente cambio de variable:

$$\left|\begin{array}{c} \frac{x_{k+1}-\mu_{k+1}}{\sigma_{k+1}} = s \\ dx_{k+1} = \sigma_{k+1}ds \end{array}\right| ,$$

obtenemos que:

$$g_{k+1}(x_1, x_{k+2}, x_{k+3}, \ldots, x_n) = 1 - \int_{-\infty}^{\infty}\frac{1}{\sqrt{2\pi}}e^{-s^2/2} \cdot \Phi\left(\frac{a_{k+1}\sigma_{k+1}s + \sum_{\substack{i=1\\ i\neq 2,\ldots,k+1}}^{n} a_i(x_i - \mu_i)}{\sqrt{\sum_{i=2}^{k} a_i^2\sigma_i^2 + \sum_{i=1}^{n} a_i^2\sigma_i^2}}\right)ds \ .$$

De nuevo usamos el resultado (33), tenemos que:

$$g_{k+1}(x_1, x_{k+2}, x_{k+3}, \ldots, x_n) = 1 - \frac{1}{\sqrt{2\pi}} \cdot$$

$$= 1 - \Phi\left(\frac{\sum_{j=1}^{n} a_j(x_j-\mu_j)}{\sqrt{\sum_{j=1}^{n} a_j^2\sigma_j^2}}\right) . \quad (29)$$

Ahora podemos calcular $h_1(x_1)$:

$$h_1(x_1) = \int_{\infty}^{\infty} f_n(x_n)\ldots\int_{\infty}^{\infty} f_k(x_k)\ldots\int_{\infty}^{\infty} f_2(x_2)dx_2\ldots dx_n,$$

llamando:

$$g_k(x_1, x_{k+1}, x_{k+2}, \ldots, x_n) = \int_{\infty}^{\infty} f_k(x_k)\ldots\int_{\infty}^{\infty} f_2(x_2)dx_2\ldots dx_k \quad (30)$$

con $k = 2, \ldots, n$, sabemos que:

$$h_1(x_1) = g_n(x_1) \ . \quad (31)$$

Vamos a demostrar por inducción sobre k que:

$$g_k(x_1, x_{k+1}, x_{k+2}, \ldots, x_n) = 1 - \Phi\left(\frac{\sum_{i\neq 2,\ldots,k}^{n}{}_{i=1}\, a_i(x_i-\mu_i)}{\sqrt{\sum_{i=2}^{k} a_i^2\sigma_i^2 + \sum_{i=1}^{n} a_i^2\sigma_i^2}}\right) . \quad (32)$$

Para ello lo demostramos primero para $k = 2$, después tomamos como hipótesis de inducción el caso k y demostramos la validez de (32) para el caso $k+1$.

Antes de demostrar la fórmula para el caso $k = 2$ es necesario introducir tres fórmulas [16] que nos serán de utilidad a la hora de hacer los cálculos:

$$I_0(a,b) = \int_{-\infty}^{\infty} e^{-s^2/2}\Phi(as+b)ds = \sqrt{2\pi}\,\Phi\left(\frac{b}{\sqrt{1+a^2}}\right) \quad (33)$$

$$I_1(a,b) = \int_{-\infty}^{\infty} s e^{-s^2/2}\Phi(as+b)ds = \frac{a}{\sqrt{1+a^2}}\exp\left(\frac{-1}{2}\frac{b^2}{1+a^2}\right) \quad (34)$$

$$I_2(a,b) = \int_{-\infty}^{\infty} s^2 e^{-s^2/2}\Phi(as+b)ds = \sqrt{2\pi}\,\Phi\left(\frac{b}{\sqrt{1+a^2}}\right) - \frac{a^2 b}{(1+a^2)\sqrt{1+a^2}}\exp\left(\frac{-1}{2}\frac{b^2}{1+a^2}\right) . \quad (35)$$

Ahora ya podemos verificar que se cumple (32) para $k = 2$:

$$g_2(x_1, x_3, x_4, \ldots, x_n) = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma_2} e^{-(x_2-\mu_2)^2/2\sigma_2^2} \cdot \left(1 - \Phi\left(\frac{\sum_{i=1}^{n} a_i(x_i-\mu_i)}{\sqrt{\sum_{i=1}^{n} a_i^2\sigma_i^2}}\right)\right) dx_2 \ .$$

Efectuando el cambio de variable:

$$\left| \begin{array}{c} \frac{x_2-\mu_2}{\sigma_2} = s \\ dx_2\sigma_2 ds \end{array} \right| ,$$

obtenemos que:

$$g_2(x_1, x_3, x_4, \ldots, x_n) = 1 - \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-s^2/2} \cdot \Phi\left(\frac{a_2\sigma_2 s + a_2\mu_2 + \sum_{i=1, i\neq 2}^{n} a_i x_i - \sum_{i=1}^{n} a_i\mu_i}{\sqrt{\sum_{i=1}^{n} a_i^2\sigma_i^2}}\right) ds$$

$$= 1 - \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{-s^2/2} \cdot \Phi\left(\frac{a_2\sigma_2 s}{\sqrt{\sum_{i=1}^{n} a_i^2\sigma_i^2}} + \frac{\sum_{i=1, i\neq 2}^{n} a_i(x_i-\mu_i)}{\sqrt{\sum_{i=1}^{n} a_i^2\sigma_i^2}}\right) ds \ .$$

Utilizando el resultado (33):

$$g_2(x_1, x_3, x_4, \ldots, x_n) = 1 - \frac{1}{\sqrt{2\pi}} \cdot I_0\left(\frac{a_2\sigma_2}{\sqrt{\sum_{i=1}^{n} a_i^2\sigma_i^2}}, \frac{\sum_{i=1, i\neq 2}^{n} a_i(x_i-\mu_i)}{\sqrt{\sum_{i=1}^{n} a_i^2\sigma_i^2}}\right)$$

$$= 1 - \Phi\left(\frac{\sum_{i=1, i\neq 2}^{n} a_i(x_i-\mu_i)}{\sqrt{a_2^2\sigma_2^2 + \sum_{i=1}^{n} a_i^2\sigma_i^2}}\right) . \quad (36)$$

Por la hipótesis de inducción suponemos que (32) es cierto para k, veamos que también es cierto para $k+1$:

$$g_{k+1}(x_1, x_{k+2}, x_{k+3}, \ldots, x_n) = \int_{-\infty}^{\infty} f_{k+1}(x_{k+1}) g_k(x_1, x_{k+1}, x_{k+2}, \ldots, x_n) dx_{k+1}$$

$$= 1 - \int_{-\infty}^{\infty} f_{k+1}(x_{k+1}) \cdot \Phi\left(\frac{\sum_{i\neq 2,\ldots,k}^{n}{}_{i=1}\, a_i(x_i-\mu_i)}{\sqrt{\sum_{i=2}^{k} a_i^2\sigma_i^2 + \sum_{i=1}^{n} a_i^2\sigma_i^2}}\right) dx_{k+1} \ .$$

Haciendo el siguiente cambio de variable:

$$\left| \begin{array}{c} \frac{x_{k+1}-\mu_{k+1}}{\sigma_{k+1}} = s \\ dx_{k+1} = \sigma_{k+1} ds \end{array} \right| ,$$

obtenemos que:

$$g_{k+1}(x_1, x_{k+2}, x_{k+3}, \ldots, x_n) = 1 - \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-s^2/2} \cdot \Phi\left(\frac{a_{k+1}\sigma_{k+1} s + \sum_{i\neq 2,\ldots,k+1}^{n}{}_{i=1}\, a_i(x_i-\mu_i)}{\sqrt{\sum_{i=2}^{k} a_i^2\sigma_i^2 + \sum_{i=1}^{n} a_i^2\sigma_i^2}}\right) ds \ .$$

De nuevo usamos el resultado (33), tenemos que:

$$g_{k+1}(x_1, x_{k+2}, x_{k+3}, \ldots, x_n) = 1 - \frac{1}{\sqrt{2\pi}} \cdot$$

$$\cdot I_0\left(\frac{a_{k+1}\sigma_{k+1}}{\sqrt{\sum_{i=2}^{k}a_i^2\sigma_i^2+\sum_{i=1}^{n}a_i^2\sigma_i^2}}\ ,\ \frac{\sum_{\substack{i=1\\ i\neq 2,\ldots,k+1}}^{n} a_i(x_i-\mu_i)}{\sqrt{\sum_{i=2}^{k}a_i^2\sigma_i^2+\sum_{i=1}^{n}a_i^2\sigma_i^2}}\right).$$

Sustituyendo el valor que corresponde a I_0 en ese punto:

$$g_{k+1}(x_1,x_{k+2},x_{k+3},\ldots,x_n)=1$$
$$-\Phi\left(\frac{\frac{\sum_{\substack{i=1\\ i\neq 2,\ldots,k+1}}^{n} a_i(x_i-\mu_i)}{\sqrt{\sum_{i=2}^{k}a_i^2\sigma_i^2+\sum_{i=1}^{n}a_i^2\sigma_i^2}}}{\sqrt{1+\frac{a_{k+1}^2\sigma_{k+1}^2}{\sum_{i=2}^{k}a_i^2\sigma_i^2+\sum_{i=1}^{n}a_i^2\sigma_i^2}}}\right)$$
$$=1-\Phi\left(\frac{\sum_{\substack{i=1\\ i\neq 2,\ldots,k+1}}^{n} a_i(x_i-\mu_i)}{\sqrt{\sum_{i=2}^{k+1}a_i^2\sigma_i^2+\sum_{i=1}^{n}a_i^2\sigma_i^2}}\right). \quad (37)$$

Por tanto hemos probado que se cumple (32) y entonces:

$$h_1(x_1)=g_n(x_1)=1-\Phi\left(\frac{a_1(x_1-\mu_1)}{\sqrt{a_1^2\sigma_1^2+2\sum_{i=2}^{n}a_i^2\sigma_i^2}}\right). \quad (38)$$

Ahora ya podemos calcular μ_1^{t+1}:

$$\mu_1^{t+1}=\int_{-\infty}^{\infty}2x_1f_1(x_1)h_1(x_1)dx_1$$
$$=\int_{-\infty}^{\infty}2\frac{1}{\sqrt{2\pi}\sigma_1}x_1e^{-(x_1-\mu_1)^2/2\sigma_1^2}\cdot$$
$$\cdot\left(1-\Phi\left(\frac{a_1(x_1-\mu_1)}{\sqrt{a_1^2\sigma_1^2+2\sum_{i=2}^{n}a_i^2\sigma_i^2}}\right)\right)dx_1\ .$$

Utilizando una nueva variable:

$$\left|\begin{array}{l}\frac{x_1-\mu_1}{\sigma_1}=s\\ dx_1=\sigma_1 ds\end{array}\right|,$$

obtenemos:

$$\mu_1^{t+1}=\frac{2}{\sqrt{2\pi}}\int_{-\infty}^{\infty}(\sigma_1 s+\mu_1)e^{-s^2/2}\cdot$$
$$\cdot\left(1-\Phi\left(\frac{a_1\sigma_1 s}{\sqrt{a_1^2\sigma_1^2+2\sum_{i=2}^{n}a_i^2\sigma_i^2}}\right)\right)ds$$
$$=\frac{2}{\sqrt{2\pi}}\left[\sigma_1\int_{-\infty}^{\infty}se^{-s^2/2}ds+\mu_1\int_{-\infty}^{\infty}e^{-s^2/2}ds\right.$$
$$-\sigma_1\int_{-\infty}^{\infty}se^{-s^2/2}\Phi\left(\frac{a_1\sigma_1 s}{\sqrt{a_1^2\sigma_1^2+2\sum_{i=2}^{n}a_i^2\sigma_i^2}}\right)ds$$
$$\left.-\mu_1\int_{-\infty}^{\infty}e^{-s^2/2}\Phi\left(\frac{a_1\sigma_1 s}{\sqrt{a_1^2\sigma_1^2+2\sum_{i=2}^{n}a_i^2\sigma_i^2}}\right)ds\right].$$

Expresando las integragrales anteriores con ayuda de las ecuaciones (33) y (34):

$$\mu_1^{t+1}=\frac{2}{\sqrt{2\pi}}\left[\sigma_1\cdot 0+\mu_1\cdot\sqrt{2\pi}\right.$$
$$-\sigma_1\cdot I_1\left(\frac{a_1\sigma_1}{\sqrt{a_1^2\sigma_1^2+2\sum_{i=2}^{n}a_i^2\sigma_i^2}}\ ,\ 0\right)$$
$$\left.-\mu_1\cdot I_0\left(\frac{a_1\sigma_1}{\sqrt{a_1^2\sigma_1^2+2\sum_{i=2}^{n}a_i^2\sigma_i^2}}\ ,\ 0\right)\right].$$

Empleando los resultados (33) y (34):

$$\mu_1^{t+1}=\frac{2}{\sqrt{2\pi}}\left[\mu_1\sqrt{2\pi}-\sigma_1\frac{a_1\sigma_1}{\sqrt{2a_1^2\sigma_1^2+2\sum_{i=2}^{n}a_i^2\sigma_i^2}}\right.$$
$$\left.-\mu_1\frac{\sqrt{2\pi}}{2}\right]=\mu_1-\frac{a_1\sigma_1^2}{\sqrt{\pi}\sqrt{\sum_{i=1}^{n}a_i^2\sigma_i^2}}\ . \quad (39)$$

Recapitulando:

$$\mu_1^{t+1}=\mu_1^t-\frac{a_1(\sigma_1^t)^2}{\sqrt{\pi}\sqrt{\sum_{i=1}^{n}a_i^2(\sigma_i^t)^2}}\ . \quad (40)$$

La expresión para la esperanza en una componente cualquiera i se obtiene de la manera análoga:

$$\mu_i^{t+1}=\mu_i^t-\frac{a_i(\sigma_i^t)^2}{\sqrt{\pi}\sqrt{\sum_{j=1}^{n}a_j^2(\sigma_j^t)^2}}\ . \quad (41)$$

B. Cálculo de σ_i^{t+1}

Al igual que en el cálculo de μ_i^{t+1} sólo realizaremos los cálculos para el caso $i=1$ y luego generalizaremos el resultado:

$$(\sigma_1^{t+1})^2=\mathsf{Var}[X_{(1:2),1}^{t+1}]$$
$$=\mathsf{E}\left[(X_{(1:2),1}^{t+1})^2\right]-\left(\mathsf{E}[X_{(1:2),1}^{t+1}]\right)^2. \quad (42)$$

Empezamos por obtener $\mathsf{E}\left[(X_{(1:2),1}^{t})^2\right]$

$$\mathsf{E}\left[(X_{(1:2),1}^{t+1})^2\right]=\int_{-\infty}^{\infty}2x_1^2f_1(x_1)h(x_1)dx_1$$
$$=\int_{-\infty}^{\infty}2\frac{1}{\sqrt{2\pi}\sigma_1}x_1^2e^{-(x_1-\mu_1)^2/2\sigma_1^2}\cdot$$
$$\cdot\left(1-\Phi\left(\frac{a_1(x_1-\mu_1)}{\sqrt{a_1^2\sigma_1^2+2\sum_{i=2}^{n}a_i^2\sigma_i^2}}\right)\right)dx_1\ .$$

Cambiando de variable:

$$\left|\begin{array}{l}\frac{x_1-\mu_1}{\sigma_1}=s\\ dx_1\sigma_1 ds\end{array}\right|,$$

tenemos que:

$$\mathsf{E}\left[(X_{(1:2),1}^{t+1})^2\right]=\frac{2}{\sqrt{2\pi}}\int_{-\infty}^{\infty}(\sigma_1 s+\mu_1)^2e^{-s^2/2}\cdot$$
$$\cdot\left(1-\Phi\left(\frac{a_1\sigma_1 s}{\sqrt{a_1^2\sigma_1^2+2\sum_{i=2}^{n}a_i^2\sigma_i^2}}\right)\right)ds$$
$$=\frac{2}{\sqrt{2\pi}}\left[\sigma_1^2\int_{-\infty}^{\infty}s^2e^{-s^2/2}ds\right.$$
$$+2\sigma_1\mu_1\int_{-\infty}^{\infty}se^{-s^2/2}ds+\mu_1^2\int_{-\infty}^{\infty}e^{-s^2/2}ds$$

$$-\sigma_1^2\int_{-\infty}^{\infty} s^2 e^{-s^2/2}\Phi\left(\frac{a_1\sigma_1 s}{\sqrt{a_1^2\sigma_1^2+2\sum_{i=2}^n a_i^2\sigma_i^2}}\right)ds$$

$$-2\sigma_1\mu_1\int_{-\infty}^{\infty} s e^{-s^2/2}\Phi\left(\frac{a_1\sigma_1 s}{\sqrt{a_1^2\sigma_1^2+2\sum_{i=2}^n a_i^2\sigma_i^2}}\right)ds$$

$$-\mu_1^2\int_{-\infty}^{\infty} e^{-s^2/2}\Phi\left(\frac{a_1\sigma_1 s}{\sqrt{a_1^2\sigma_1^2+2\sum_{i=2}^n a_i^2\sigma_i^2}}\right)ds\Bigg].$$

Escribiendo las integrales anteriores en función de (33), (34) y (35)podemos decir que:

$$\mathsf{E}\left[(X^{t+1}_{(1:2),1})^2\right]=\frac{2}{\sqrt{2\pi}}\Bigg[\sigma_1^2\cdot\sqrt{2\pi}+2\sigma_1\mu_1\cdot 0$$

$$+\mu_1^2\cdot\sqrt{2\pi}-\sigma_1^2\cdot I_2\left(\frac{a_1\sigma_1}{\sqrt{a_1^2\sigma_1^2+2\sum_{i=2}^n a_i^2\sigma_i^2}},\,0\right)$$

$$-2\sigma_1\mu_1\cdot I_1\left(\frac{a_1\sigma_1}{\sqrt{a_1^2\sigma_1^2+2\sum_{i=2}^n a_i^2\sigma_i^2}},\,0\right)$$

$$-\mu_1^2\cdot I_0\left(\frac{a_1\sigma_1}{\sqrt{a_1^2\sigma_1^2+2\sum_{i=2}^n a_i^2\sigma_i^2}},\,0\right)\Bigg].$$

Utilizando (33), (34) y (35) obtenemos:

$$\mathsf{E}\left[(X^{t+1}_{(1:2),1})^2\right]=\frac{2}{\sqrt{2\pi}}\Bigg[\sigma_1^2\sqrt{2\pi}+\mu_1^2\sqrt{2\pi}$$

$$-\sigma_1^2\frac{\sqrt{2\pi}}{2}-2\sigma_1\mu_1\frac{a_1\sigma_1}{\sqrt{2\sum_{i=1}^n a_i^2\sigma_i^2}}-\mu_1^2\frac{\sqrt{2\pi}}{2}\Bigg]$$

$$=\sigma_1^2+\mu_1^2-\frac{2a_1\mu_1\sigma_1^2}{\sqrt{\pi}\sqrt{\sum_{i=1}^n a_i^2\sigma_i^2}}\;. \tag{43}$$

Escribimos los superíndices correspondientes al paso para obtener la expresión completa:

$$\mathsf{E}\left[(X^{t+1}_{(1:2),1})^2\right]=(\sigma_1^t)^2+(\mu_1^t)^2-\frac{2a_1\mu_1^t(\sigma_1^t)^2}{\sqrt{\pi}\sqrt{\sum_{i=1}^n a_i^2(\sigma_i^t)^2}}. \tag{44}$$

Finalmente:

$$(\sigma_1^{t+1})^2=(\sigma_1^t)^2\left(1-\frac{a_1^2(\sigma_1^t)^2}{\pi\sum_{i=1}^n a_i^2(\sigma_i^t)^2}\right),$$

equivalentemente:

$$\sigma_1^{t+1}=\sigma_1^t\sqrt{1-\frac{a_1^2(\sigma_1^t)^2}{\pi\sum_{i=1}^n a_i^2(\sigma_i^t)^2}}\;. \tag{45}$$

Se puede escribir una expresión análoga para una componente cualquiera i :

$$\sigma_i^{t+1}=\sigma_i^t\sqrt{1-\frac{a_i^2(\sigma_i^t)^2}{\pi\sum_{j=1}^n a_j^2(\sigma_j^t)^2}}\;. \tag{46}$$

V. Comportamiento del algoritmo

Una vez obtenidas las expresiones de μ_i^{t+1} y σ_i^{t+1} el siguiente paso sería intentar predecir el comportamiento del algoritmo cuando t aumenta. Ésto se consigue analizando cada sucesión de medias $\{\mu_i^t\}_t$ y cada sucesión de varianzas $\{\sigma_i^t\}_t$ con $t\in I\!N$.

Para probar que el algoritmo funciona de forma adecuada deberíamos ver que:

$$\text{cuando}\begin{cases} a_i>0\Rightarrow\mu_i^t\longrightarrow_{t\to\infty}\infty\\ a_i<0\Rightarrow\mu_i^t\longrightarrow_{t\to\infty}-\infty\end{cases}, \tag{47}$$

puesto que en ese caso el algoritmo mejoraría a cada paso, minimizando el valor de función objetivo indefinidamente.

Desgraciadamente las sucesiones de medias $\{\mu_i^t\}_t$ con $t\in I\!N$ son difíciles de estudiar cuando las σ_i^t son distintas. Aún así, podemos decir que la mejora a cada paso y en cada componente viene dada por:

$$|\mu_i^{t+1}-\mu_i^t|=\left|\frac{a_i(\sigma_i^t)^2}{\sqrt{\pi}\sqrt{\sum_{j=1}^n a_j^2(\sigma_j^t)^2}}\right| \tag{48}$$

Por lo tanto, dado que las sucesiónes $\{\sigma_i^t\}_t$ son decrecientes para todo i, la mejora en cada componente va disminuyendo según aumenta t.

Dada la dificultad para analizar las sucesiones $\{\mu_i^t\}_t$ y $\{\sigma_i^t\}_t$ con $t\in I\!N$ vamos a centrarnos en un caso particular. Analizaremos la función $G(\mathbf{x})=\sum_{i=1}^n x_i$, cuando la sucesión de varianzas cumple la condición:

$$\sigma_i^t=\sigma^t\;,\text{ con } i=1,\ldots,n\;. \tag{49}$$

Estudiamos primero la sucesión de varianzas $\{\sigma^t\}_t$ con $t\in I\!N$. Dado que :

$$\sigma^{t+1}=\sigma^t\left(\frac{n\pi-1}{n\pi}\right)^{1/2}, \tag{50}$$

podemos escribir σ^{t+1} en función de σ^0. De este modo la sucesión de las varianzas se puede escribir como la siguiente sucesión recurrente:

$$\sigma^{t+1}=\sigma^0\left(\frac{n\pi-1}{n\pi}\right)^{t/2} \tag{51}$$

La expresión anterior nos ayudará en el análisis de la sucesión de medias $\{\mu^t\}_t$ con $t\in I\!N$, puesto que, sustituyéndola en la expresión de μ^t obtenemos:

$$\mu^t=\mu^{t-1}-\frac{\sigma^0}{\sqrt{n\pi}}\left(\frac{n\pi-1}{n\pi}\right)^{\frac{t-1}{2}}. \tag{52}$$

Podemos además expresar μ^t en función de μ^0 y de σ^0:

$$\mu^t=\mu^0-\frac{\sigma^0}{\sqrt{n\pi}}\left(1+\left(\frac{n\pi-1}{n\pi}\right)^{1/2}+\ldots+\left(\frac{n\pi-1}{n\pi}\right)^{\frac{t-1}{2}}\right)$$

operando podemos ver que la sucesión de las medias es también una sucesión recurrente:

$$\mu^t=\mu^0-\frac{\sigma^0}{\sqrt{n\pi}}\cdot\frac{\left(\frac{n\pi-1}{n\pi}\right)^{t/2}-1}{\left(\frac{n\pi-1}{n\pi}\right)^{1/2}-1}. \tag{53}$$

Una vez encontrada la expresión anterior la sucesión de medias es más sencilla de analizar. Ésta sucesión es decreciente y además tiene límite finito:

$$\begin{aligned}\lim_{t\to\infty}\mu^t &= \lim_{t\to\infty}\left(\mu^0 - \frac{\sigma^0}{\sqrt{n\pi}}\cdot\frac{\left(\frac{n\pi-1}{n\pi}\right)^{\frac{t}{2}}-1}{\left(\frac{n\pi-1}{n\pi}\right)^{1/2}-1}\right)\\ &= \mu^0 + \frac{\sigma^0}{\sqrt{n\pi}}\frac{1}{\left(\frac{n\pi-1}{n\pi}\right)^{1/2}-1}\\ &= \mu^0 + \frac{\sigma^0}{\sqrt{n\pi-1}-\sqrt{n\pi}}\ . \end{aligned} \tag{54}$$

El hecho de que la sucesión de medias tenga límite finito implica un mal comportamiento en términos del algoritmo, ya que, como hemos dicho anteriormente un buen funcionamiento del algoritmo se corresponde con un decrecimiento de μ^t indefinido y no acotado.

Por lo tanto podemos concluir que éste algoritmo no tiene un buen comportamiento cuando se realiza una cantidad muy grande de torneos a cada paso.

VI. Agradecimientos

Queremos agradecer la inestimable colaboración del profesor J. de Cal.

El presente trabajo ha sido realizado dentro del proyecto de investigación 9/UPV/EHU 00140.226-12084/2000. Además C. González disfruta una beca predoctoral concedida por la Universidad del País Vasco - Euskal Herriko Unibertsitatea.

VII. Conclusiones

Dos han sido los logros en éste trabajo. Por un lado se ha conseguido modelar matemáticamente el algoritmo UMDA$_c$ con selección por torneo aplicado a funciones lineales y realizando un número infinito de torneos a cada paso. Por otro, se ha analizado el comportamiento del mismo, basándose en el modelo matemático, llegando a la conclusión de que el algoritmo no tiene un buen comportamiento en la función lineal $G_1(\mathbf{x}) = \sum_{i=1}^n x_i$, con $\mathbf{x} = (x_1, \ldots, x_n) \in I\!R^n$ cuando se realiza un número muy elevado de torneos a cada paso.

A partir de ahora nuestro interés se centrará en el estudio del algoritmo aplicado a otras funciones, como $G(\mathbf{x}) = \sum_{i=1}^n x_i^2$ ó $G(\mathbf{x}) = \sum_{i=1}^n |x_i|$. También intentaremos hacer un estudio análogo para la versión discreta del algoritmo.

Referencias

[1] P. Larrañaga and J. A. Lozano, *Estimation of Distribution Algorithms. A New Tool for Evolutionary Computation*, Kluwer Academic Publishers, 2001.

[2] H. Mühlenbein and G. Paaβ, "From Recombination of Genes to the Estimation of Distributions I. Binary Parameters," in *Parallel Problem Solving from Nature, PPSN-IV*, H.-M. Voigt, W. Ebeling, I. Rechenberg, and H.-P. Schwefel, Eds., 1996, pp. 178–187.

[3] P. Larrañaga, R. Etxeberria, J. A. Lozano, and J. M. Peña, "Combinatorial Optimization by Learning and Simulation of Bayesian Networks," in *Proceedings of Uncertainty in Artificial Intelligence, UAI-2000*, C. Boutilier and M. Goldszmidt, Eds. 2000, pp. 343–352, Morgan Kaufmann.

[4] P. Larrañaga, R. Etxeberria, J. A. Lozano, and J. M. Peña, "Optimization in Continuous Domains by Learning and Simulation of Gaussian Networks," in *Proceedings of Genetic and Evolutionary Computation Conference, Workshop Program*, A. S. Wu, Ed., 2000, pp. 201–204.

[5] M. Pelikan, D. E. Goldberg, and F. Lobo, "A Survey of Optimization by Building and Using Probabilistic Models," Tech. Rep. IlliGAL, 99018, Urbana, IL: University of Illinois at Urbana-Champaing, Illinois Genetic Algorithms Laboratory, 1999.

[6] P. Larrañaga, R. Etxeberria, J. A. Lozano, and J. M. Peña, "Optimization by Learning and Simulation of Bayesian and Gaussian Networks," Tech. Rep. KZZA-IK-4-99, University of the Basque Country, Department of Computer Science and Artificial Intelligence, University of the Basque Country, 1999.

[7] C. González, J. A. Lozano, and P. Larrañaga, "Mathematical Modelling of Discrete Estimation of Distribution Algorithms," in *Estimation of Distribution Algorithms. A New Tool for Evolutionary Computation*, P. Larrañaga and J. A. Lozano, Eds. 2001, Kluwer Academic Publishers.

[8] H. Mühlenbein, T. Mahnig, and A. Ochoa, "Schemata, Distributions and Graphical Models in Evolutionary Optimization," *Journal of Heuristics*, vol. 5, pp. 215–247, 1999.

[9] H. Mühlenbein and T. Mahnig, "Application of Wright's Equation in Evolutionary Computation," in *Proceedings of Advances in Fuzzy Systems and Evolutionary Computation, WSES*, 2001, pp. 265–270.

[10] M. Höhfeld and G. Rudolph, "Towards a Theory of Population-Based Incremental Learning," in *Proceedings of the Fourth IEEE Conference on Evolutionary Computation*. 1997, IEEE Press.

[11] C. González, J. A. Lozano, and P. Larrañaga, "The Convergence Behavior of the PBIL Algorithm: A Preliminary Approach," in *Proceedings of the Fifth International Conference on Artificial Neural Networks and Genetic Algorithms, ICANNGA'2001*, V Kurková, N. C. Steele, R. Neruda, and M. Kárný, Eds., 2001, pp. 228–231.

[12] C. González, J. A. Lozano, and P. Larrañaga, "Analyzing the Population Based Incremental Learning Algorithm by Means of Discrete Dynamical Systems," *Complex Systems*, vol. 12, no. 4, pp. 465–479, 2000.

[13] A. Berny, "An Adaptative Scheme for Real Function Optimization Acting as a Selection Operator," in *Proceedings of First IEEE Symposium on Combinations of Evolutionary Computation and Neural Networks*, X. Yao, Ed., 2000.

[14] A. Berny, "Selection and Reinforcement Learning for Combinatorial Optimization," in *Proceedings of Parallel Problem Solving from Nature, PPSN-VI*, M. Schoenauer, K. Deb, G. Rudolph, X. Yao, E. Lutton, J. J. Merelo, and H.-P. Schwefel, Eds. 2000, Springer-Verlag, Berlin Heidelberg.

[15] H. Mühlenbein, "The Equation for Response to Selection an its Use for Prediction," *Evolutionary Computation*, vol. 5, pp. 303–346, 1998.

[16] H. G. Beyer, *The Theory of Evolution Strategies*, Springer, 2001.