



Centro de Investigación en Matemáticas, A.C.

**Una propuesta para
Algoritmos de Estimación de
Distribución no paramétricos**

T E S I S

Que para obtener el grado de

Maestro en Ciencias

con Orientación en

Computación y Matemáticas Industriales

P R E S E N T A :

Selma Ela Barba Moreno

Director de Tesis:

Dr. Arturo Hernández Aguirre

A mis padres

Agradecimientos

Ante todo quisiera expresar mi gratitud al director de la tesis, el Dr. Arturo Hernández Aguirre por su tiempo, su dedicación y sus valiosos consejos. Así mismo, me gustaría agradecer al coasesor de este trabajo, el Dr. Enrique Raúl Villa Diharce por el gran apoyo y el inestimable tiempo ofrecido para formalizar los conceptos revisados en esta investigación.

También agradezco a los revisores de este documento, el Dr. Salvador Botello Rionda y el Dr. Johan Jozef Lode Van Horebeek, por sus observaciones, comentarios y sugerencias, las cuales ayudaron a mejorar la calidad de este trabajo de investigación.

Igualmente, quisiera extender mi gratitud al Consejo Nacional de Ciencia y Tecnología (CONACyT) por haber aportado el financiamiento necesario para la realización de mis estudios de maestría, los cuales culminan con la elaboración del presente trabajo.

Agradezco enormemente a mi familia por el apoyo ofrecido, por haber creído en mí y haberme impulsado para seguir adelante, a mi padre por su enseñanza moral y a mi madre por su amor y cuidado. Además, agradezco a William por estar conmigo en esta etapa de mi vida, por avivar mis ganas de seguir adelante y cuidarme cuando más lo necesitaba y porque nunca me dejó sola para alcanzar este sueño.

Índice general

Introducción	11
1. Elementos de la computación evolutiva	15
1.1. Conceptos básicos	16
1.2. Algoritmos Genéticos	18
1.2.1. Operadores de selección	20
1.2.2. Operadores de cruza	22
1.2.3. Operadores de mutación	23
1.3. Bloques constructores y decepción	24
1.3.1. Los problemas de los AG: las funciones deceptivas	25
1.4. Conclusiones del capítulo 1	26
2. Algoritmos con Estimación de Distribución	27
2.1. Conceptos de probabilidad	28
2.2. Algoritmos de Estimación de Distribución	31
2.2.1. EDAs con representación discreta	32
2.2.2. EDAs con representación real	36
2.3. Conclusiones del capítulo 2	39
3. Algoritmo de Estimación de Distribución no paramétrico	41
3.1. Medidas de asociación	41
3.2. Inducción de la correlación	44
3.3. Descripción del algoritmo	47
3.4. Conclusiones del capítulo 3	52
4. Algoritmo de Estimación de Distribución con Cópulas	53
4.1. Cópulas: definiciones y propiedades.	54
4.1.1. Familias paramétricas de cópulas	57
4.2. Medidas de asociación	62
4.2.1. Coeficiente de correlación por rangos de Kendall	62
4.2.2. Coeficiente de correlación por rangos de Spearman	63
4.2.3. Relaciones entre ρ_S y τ de Kendall	64
4.2.4. Rango de las medidas de asociación	65
4.3. Construcción del modelo de probabilidad usando cópulas	66

4.3.1. Simulación condicional	66
4.3.2. Simulación elíptica	68
4.4. Algoritmo de Estimación de Distribución con Cópulas: COPU-LEDA	73
4.5. Ajuste de la varianza	75
4.5.1. Transformación lineal	76
4.5.2. Transformación no lineal	77
4.5.3. Tiempo de ajuste	80
4.6. Conclusiones del capítulo 4	81
5. Experimentos y resultados	83
5.1. Conjuntos experimentales	83
5.2. Resultados	87
5.2.1. NOPREDA sin ajuste de varianza	88
5.2.2. NOPREDA con ajuste de varianza	91
5.2.3. COPULEDA sin ajuste de varianza	94
5.2.4. COPULEDA con ajuste de varianza	97
5.3. Comparación estadística del comportamiento de los algoritmos .	100
5.3.1. Resultados	102
5.4. Conclusiones del capítulo 5	108
6. Conclusiones generales	111
Referencias	115
A. Forma paramétrica de ρ_s en la familia Clayton	119

Índice de figuras

1.1.	La función <i>sphere</i> en tres dimensiones.	17
1.2.	Cruza de un punto	19
1.3.	Cruza de cuatro puntos	22
1.4.	Mutación de un individuo con representación binaria	23
2.1.	Clasificación de los EDAs.	32
2.2.	Representación gráfica del BMDA.	35
2.3.	Representación gráfica del PADA.	36
3.1.	Interpolación de los datos para obtener $\hat{F}(x)$	48
3.2.	Frecuencia acumulada empírica.	49
3.3.	Muestreo aleatorio para regenerar la población.	49
3.4.	Población inicial	50
3.5.	Población regenerada mediante Normales Multivariadas	51
3.6.	Población regenerada mediante NOPREDA	51
4.1.	Función de distribución conjunta $H(x, y)$	57
4.2.	Curvas de nivel de una cópula normal.	58
4.3.	Curvas de nivel de la cópula Clayton.	61
4.4.	Curvas de nivel de la cópula Frank.	61
4.5.	Curvas de nivel de la cópula Gumbel.	61
4.6.	Comportamiento de ρ_S y τ en variables continuas.	64
4.7.	Comportamiento de τ en la familia FGM.	66
4.8.	Comportamiento de τ en la familia Normal.	66
4.9.	Comportamiento de τ en la familia Clayton.	66
4.10.	Comportamiento de τ en la familia Frank.	66
4.11.	Función $H(x, y)$ obtenida con una cópula Normal.	70
4.12.	Función $H(x, y)$ obtenida con normales independientes.	71
4.13.	Función $H(x, y)$ obtenida con NOPREDA.	71
4.14.	Gráfica C_1^t contra C_2^t de los datos originales.	72
4.15.	Gráfica C_1^{t+1} vs C_2^{t+1} de los datos con NOPREDA	72
4.16.	Gráfica C_1^{t+1} vs C_2^{t+1} de los datos con normales independientes.	73
4.17.	Gráfica C_1^{t+1} vs C_2^{t+1} de los datos con cópula Normal.	73
4.18.	Obtención de los datos uniformes.	74

Índice de figuras

4.19. Población regenerada mediante COPULEDA	75
4.20. Ajuste de la varianza de una variable aleatoria X	76
4.21. Ajuste de varianza de una conjunto de datos.	78
4.22. Gráfica de los datos originales y los transformados con $F_{\text{máx}} = 2$	80
5.1. Función $f_{A,1}$	84
5.2. Función $f_{A,8}$	84
5.3. Función $f_{B,1}$	86
5.4. Función $f_{B,2}$	86
5.5. Función $f_{B,3}$	86
5.6. Función $f_{B,6}$	87
5.7. Función $f_{B,7}$	87
5.8. Función $f_{C,1}$	87
5.9. Función $f_{C,3}$	87

Introducción

Desde sus inicios el hombre ha buscado la forma de controlar su entorno, entender el comportamiento de los fenómenos naturales y poder realizar predicciones. Su deseo por controlar y someter lo ha llevado a la búsqueda de herramientas que le permitan optimizar: encontrar la ruta más corta, almacenar la mayor cantidad de alimento, minimizar los costos, etc., y ha encontrado algunas técnicas matemáticas que le ayudan en su propósito, asegurando el máximo beneficio a sus intereses. Sin embargo, no todas las representaciones de los fenómenos naturales pueden resolverse eficientemente empleando esta clase de técnicas e incluso hay algunas cuya solución sólo puede encontrarse mediante métodos de búsqueda aleatoria. Los algoritmos genéticos, por ejemplo, tienen características propias de la naturaleza, que los hacen instrumentos útiles para la optimización de funciones cuyo costo de cómputo es alto con otras técnicas de optimización numérica.

Estos algoritmos nacen como una forma de emular el comportamiento de los entes biológicos durante su proceso de evolución, utilizando herramientas computacionales y matemáticas. Estos algoritmos emplean elementos a nivel de genes para formar a sus individuos y realizar los avances evolutivos que permiten concebir especies mejor adaptadas. La forma general de un algoritmo genético define al *individuo* como un organismo viviente, el cual se desarrolla en una *población* gracias a la cual se mejora la especie.

Un individuo se compone de cromosomas y éstos a su vez en genes, los cuales, dada su calidad, dictan el valor del individuo y su permanencia en las generaciones futuras. Cada individuo que conforma la población es una solución potencial al problema de optimización, y aquel que presente una mayor calidad o *aptitud* representa un individuo con una solución más próxima al óptimo o incluso óptima.

Los componentes principales de la teoría neo-Darwinista se incluyen en los algoritmos genéticos y son conocidos como *función de aptitud* y operadores de *selección*, *cruza* y *mutación*. En la selección, los mejores individuos (aquellos con una función de aptitud mayor) son elegidos para recombinarse de tal suerte que su material genético prevalezca en las futuras generaciones, mientras que la mutación se refiere a pequeños cambios a nivel de genes, que permiten variaciones en el código genético de los individuos.

Estos procesos se realizan durante cientos o miles de generaciones, tiempo en el cual, la población avanza hacia jerarquías más altas en la cadena evolutiva

hasta que se encuentra al individuo que cumple las condiciones de optimalidad y entonces la evolución termina.

La propuesta de este tipo de algoritmos, desde sus inicios en la década de los 60's, ha dirigido la atención hacia los modelos de búsqueda estocástica, así como por las nuevas versiones que permiten disminuir el número de generaciones para encontrar al óptimo y hacerlo con la mayor precisión posible. Muchas de las pruebas realizadas en la actualidad han demostrado la eficacia de estos algoritmos en una gran variedad de problemas de optimización con representación discreta y continua tanto en problemas globales como con restricciones; sin embargo, el diseño de algoritmos más poderosos ha guiado el desarrollo de funciones cuya principal finalidad es la de *engañar* al algoritmo genético. Estas funciones son conocidas como *funciones deceptivas* y muestran que los algoritmos genéticos carecen de la capacidad para identificar las relaciones que existen entre los cromosomas y destruyen los elementos compositivos que, a lo largo de las generaciones, podrían refinarse en una solución factible del problema de optimización.

Es así como surge un segundo tipo de algoritmos evolutivos conocidos como *algoritmos de estimación de distribución* (EDAs por sus siglas en inglés) los cuales siguen considerando a cada solución potencial como a un individuo, pero carecen de los operadores de cruce y mutación para obtener a los nuevos organismos. En lugar de esto, toman cada cromosoma i del individuo j , como el i -ésimo valor observado de la variable aleatoria j y modelan (o aproximan) la distribución conjunta de los individuos seleccionados, para generar la nueva población con igual o similar distribución.

Los algoritmos de estimación de distribución son relativamente nuevos y es por ello que las aportaciones realizadas hasta la fecha son mínimas y se mantienen como un área abierta para la investigación. En la actualidad, existen EDAs que trabajan con representación discreta, lo cual merma la precisión en los problemas de naturaleza continua y existen unos pocos que se enfocan en problemas continuos pero que suponen independencia entre las variables. Otros tantos, consideran dependencias entre las variables pero de orden bajo o bien, ajustan a la población a cierta distribución preestablecida. La desventaja de suponer esto es que en general, las variables no son independientes sino que muestran ciertas asociaciones y por otro lado, carecen de una distribución conocida o única.

La principal finalidad de este trabajo de investigación es realizar un estudio sobre algunas de las técnicas existentes para capturar todas las dependencias de un conjunto de datos e inducir las a un nuevo conjunto. Estas técnicas se caracterizan por ser libres de distribución y se aplican a problemas con representación real, lo cual presenta dos grandes ventajas: la primera es que aplicadas a los EDAs, permiten que se desempeñen en el espacio real, lo cual es de suma importancia ya que en la actualidad la mayoría de los algoritmos con estimación de distribución propuestos trabajan en el espacio discreto y esto disminuye considerablemente la precisión de los resultados, y la segunda es que al capturar

todas las dependencias de las variables sin suponer distribuciones específicas, se garantiza que los EDAs aproximen de la mejor manera posible la distribución de los datos originales.

La inclusión de estas técnicas en los algoritmos de estimación de distribución, da pie al diseño de dos propuestas que son mostradas en este trabajo de investigación y que requieren de la mención de varios puntos importantes para su base teórica.

El primero de ellos es considerar las propiedades conocidas de las distribuciones elípticas. Estas distribuciones se caracterizan por ser completamente determinadas por su media y su matriz de covarianzas, con lo cual, es posible obtener un conjunto de variables aleatorias distribuídas de forma elíptica \mathbf{X} con vector de medias μ y matriz de covarianzas Σ mediante:

$$\mathbf{X} =_d \mu + A\mathbf{Y},$$

donde A es una matriz triangular inferior la cual cumple $\Sigma = AA^t$ y \mathbf{Y} es un vector de variables aleatorias distribuídas de forma esférica. El símbolo $=_d$ indica que la equivalencia es en distribución.

Una de las distribuciones elípticas mejor conocida es la distribución Normal ($\mathbf{X} \sim \mathcal{N}(\mu, \Sigma)$), mientras que en el caso de la distribución esférica se debe emplear aquella con parámetros de independencia ($\mathbf{Y} \sim \mathcal{N}(0, I)$). Ahora bien, como en el caso de las distribuciones elípticas, se puede obtener un vector transformado \mathbf{X} que tenga una matriz de correlaciones deseadas C , empleando $\mathbf{X} = \mathbf{Y}B^t$ con $C = BB^t$. En este caso, C es la matriz de correlaciones por rangos ρ_S , ya que esta medida es más significativa en el modelado de situaciones donde las variables están relacionadas de forma monótona y no necesariamente distribuidas de forma normal.

En segunda instancia, luego de obtener el conjunto de datos relacionados de forma deseada, se procede a inducir dichas relaciones al grupo de individuos generados de la muestra. La inducción de la correlación está basada en una técnica propuesta por Iman y Conover en 1982, la cual emplea una matriz que guarda las posiciones a las que hay que reordenar una población para mantener las asociaciones requeridas entre las variables. Esta sencilla idea es la base del NOPREDA, la cual es explorada con mayor detenimiento en el capítulo 3.

El segundo algoritmo basa su funcionamiento en el concepto de cópula. El término *cópula* se refiere a una función de distribución la cual es capaz de asociar una función de distribución multivariada, con sus distribuciones marginales de bajo orden. Dicha relación fue definida por primera vez en 1959 por A. Sklar y es conocida como el *teorema de Sklar*. Desde su presentación y hasta la fecha, el interés por las diversas familias de cópulas y sus propiedades ha ido en aumento y ha motivado una gran cantidad de aplicaciones en las ciencias económicas y actuariales, entre otras. El teorema de Sklar establece:

$$F(x_1, \dots, x_n) = C(F_1(x_1), \dots, F_n(x_n); \alpha),$$

donde $C(u_1, \dots, u_n; \alpha)$ es la cópula asociada a \mathbf{X} y α es el parámetro de la cópula conocido como *parámetro de dependencia*.

En este trabajo de investigación se revisaron las principales cópulas conocidas y a partir de sus propiedades se eligió la mejor para ser aplicada en el algoritmo de estimación de distribución. Esta cópula se caracteriza porque permite medir las dependencias en todo el rango de valores, ya sean negativas o positivas (es decir, es completa), es posible extenderla al caso multivariado y es de fácil implementación. Además, empleando el concepto de cópula se pueden redefinir algunas de las medidas de asociación como son la τ de Kendall y la ρ_s de Spearman, las cuales permiten hacer un enlace entre las relaciones multivariadas que son capturadas de la muestra y la cópula que generará a las nuevas marginales.

Sobre esta misma línea se incluyeron variaciones en los algoritmos para proveerles mayor dispersión a los datos, dado que en su forma nativa, los EDAs se enfrentan a problemas de convergencia prematura. Estas nuevas vertientes se probaron a la par de sus correspondientes originales, empleando funciones de optimización global unimodales y multimodales en diversas dimensiones. Los mejores algoritmos de cada versión (ya fuera con ajuste de varianza o sin él) se compararon entre sí y se realizó una prueba *bootstrap* no paramétrica para medir el desempeño de ambos algoritmos. En el capítulo 5 se pueden encontrar todos los pormenores de estas pruebas.

Finalmente en el capítulo 6 se muestra un resumen de los temas más importantes y las conclusiones generales a las que se llegaron con este trabajo.

Capítulo 1

Elementos de la computación evolutiva

La simulación de la evolución tal y como ocurre en la naturaleza ha sido una herramienta en el ámbito de la optimización para dar pauta a la computación evolutiva. En ésta, los individuos son seleccionados, recombinados y mutados para generar nuevos entes, con capacidades superiores a sus ancestros. Dichas capacidades se miden en términos de la “calidad” del material genético que cada individuo posee y que será transmitido a los individuos de las nuevas generaciones.

Existen formas que han sido empleadas desde la década de los 60's para simular el comportamiento de la naturaleza mediante herramientas matemáticas y computacionales. En éstas los individuos son representados mediante cadenas de números las cuales son evaluadas con algún criterio establecido que indica cuál es su valía en la población. En las últimas dos décadas la comunidad de computación científica ha empleado de forma extensiva el uso de estas herramientas para resolver problemas de optimización de funciones.

Los Algoritmos Genéticos (GA), como se les denomina a estos métodos de optimización, trabajan en el ámbito real y discreto, y su investigación ha sido tan variada, que métodos tan poderosos como las Estrategias Evolutivas o la Evolución Diferencial, generan resultados que aún son difíciles de igualar con otras técnicas.

Los Algoritmos Genéticos fueron la base para modelos más complejos que trabajan con la distribución de los individuos para inducir la tendencia de la población hacia regiones donde las probabilidades de encontrar el óptimo aumentan. Dichos modelos son conocidos como Algoritmos de Estimación de Distribución (EDA por sus siglas en inglés) y los cuales debido a su reciente propuesta aún se tienen como un área abierta para la investigación.

En EDAs, al igual que su antecesor el Algoritmo Genético, los problemas pueden ser discretos o continuos, aunque una gran parte de estos modelos se ha inclinado por la parte discreta y la parte continua ha quedado relegada a tan sólo unos pocos de ellos.

El primer capítulo se encuentra organizado de la siguiente forma: primeramente se hablará sobre algunos conceptos básicos de los algoritmos evolutivos, cómo se representa el espacio de búsqueda, los individuos, la población y la función de aptitud, entre otros; en la siguiente sección se describirá el Algoritmo Genético en su forma general y su forma de implementación para posteriormente mostrar algunas de sus deficiencias y lo que dió pauta a la exploración de otras técnicas para optimizar funciones. Finalmente, se muestran los puntos más importantes del presente capítulo y se realizan las conclusiones.

1.1. Conceptos básicos

En el ámbito de los Algoritmos Evolutivos se emplea una gran cantidad de términos biológicos que es preciso definir. Estos términos son empleados de igual forma que en el contexto de la biología real, sólo que las entidades a las que se refieren estos modelos son mucha más sencillas que las biológicas.

En el caso de los seres vivos la colección del material genético se conoce como el *genotipo* del organismo, y comprende un conjunto particular de genes contenido en un genoma. El genotipo de un individuo es una forma codificada de representar al *fenotipo*, que son las características mentales y físicas del individuo.

Como en la naturaleza, los individuos de un Algoritmo Evolutivo poseen ambas componentes. El genotipo, que son los valores que toman las variables objetivo de las que están formados los integrantes de la población; y el fenotipo, que es la manifestación de ese individuo, la cual se toma como una “puntuación” que define su calidad en la población y se conoce como *aptitud*.

En términos formales, un problema de optimización se encuentra comprendido por dos elementos, uno de ellos es el espacio de búsqueda Ω y el otro es la función de optimización f , en la cual

$$f : \Omega \rightarrow \mathbb{R}. \tag{1.1}$$

El problema consiste en encontrar el punto óptimo x^* dentro de la región Ω tal que se minimice o maximice la función f si lo que se desea es minimizarla o maximizarla, respectivamente. Con lo cual el punto x^* es aquel en el que se cumple que:

$$\forall x \in \Omega : f(x^*) \geq f(x), \text{ para maximizar } f, \quad (1.2)$$

o bien

$$\forall x \in \Omega : f(x^*) \leq f(x), \text{ para minimizar } f \quad (1.3)$$

En el campo de los algoritmos evolutivos la función f es conocida como la *función de aptitud* del individuo x y define su valor en la población. La *población* es un conjunto de puntos o posibles soluciones de un problema $x \in \Omega$ conocidas como individuos o cromosomas, los cuales compiten a lo largo de las generaciones para mejorar su función de aptitud y alcanzar el punto óptimo x^* .

Como es de suponer, el problema de optimización puede ser global o contar con restricciones que acoten la región Ω , además el tipo de problema a resolver puede contener un sólo mínimo o un sinnúmero de mínimos locales y restricciones de diversos grados de dificultad.

A manera de ilustración, considérese el siguiente ejemplo. Sea,

$$f(x) = \sum_{i=1}^n x_i^2 \quad (1.4)$$

un problema de optimización, en el cual f es la función a minimizar. Observe que el problema no tiene restricciones y cuenta con un único mínimo global. La gráfica de f se puede apreciar en la figura 1.1.

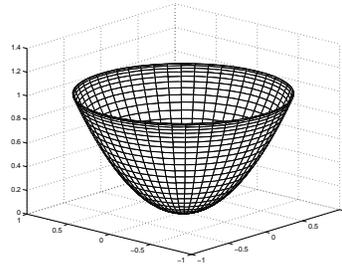


Figura 1.1: La función (1.4) en tres dimensiones.

Para este ejemplo, la región de búsqueda es $\Omega = \mathbb{R}^n$ y un punto en este espacio está representado como $x = (x_1, x_2, \dots, x_n)$, con n como la dimensión del espacio. En este caso los individuos que se aproximen al óptimo deberán contar con mejores valores de aptitud, de tal suerte que prevalezcan en la búsqueda del mejor.

Al igual que en la naturaleza, en un Algoritmo Evolutivo los individuos realizan *recombinaciones* o cruza en las que cada uno de los padres proporciona parte de su material genético a sus descendientes. Este proceso se simula mediante la combinación de los elementos de cada cadena de dos individuos padres para formar uno o varios individuos hijos. Además, los individuos se encuentran sujetos a *mutaciones*, las cuales son ligeros cambios del código genético que es proporcionado de padre a hijo. En el caso computacional los cambios se efectúan tomando en cuenta una probabilidad de mutación que permite realizar un ligero cambio en el valor de las variables del individuo con lo cual se obtiene un nuevo ente levemente diferente a su antecesor. En la siguiente sección se hablará más profundamente de estos procesos.

1.2. Algoritmos Genéticos

Los Algoritmos Genéticos (GAs), como se ha mencionado antes, tratan de semejar el comportamiento de las especies naturales, empleando a su vez, los elementos y la nomenclatura que éstos utilizan. Los Algoritmos Genéticos cuentan con los siguientes componentes básicos: una población de individuos, un conjunto de operadores aleatorios que modifican a cada individuo y un mecanismo de selección de los individuos. En el modelo más simple los individuos son cadenas de longitud n y se encuentran en el espacio $\Omega = \mathbb{B}^n = \{0, 1\}^n$, es decir que únicamente toman valores de 0 y 1.

El algoritmo 1 muestra el pseudocódigo de un Algoritmo Genético en su forma básica.

Algoritmo 1 GA

```

1: t=0
2: Inicializar  $P(t)$ .
3: Evaluar cada uno de los individuos de  $P(t)$ .
4: while no se cumpla la condición de paro do
5:   for  $i = 1$  to  $N/2$  do
6:     Seleccionar dos padres de  $P(t)$ .
7:     Aplicar la cruza a los dos padres con probabilidad  $p_C$ .
8:     Mutar los hijos con probabilidad  $p_M$ .
9:     Generar la nueva población con los hijos generados.
10:  end for
11:  t=t+1
12: end while

```

En este algoritmo, N representa el tamaño de la población y $P(t)$ es la población en la generación t . Observe que el criterio de paro puede ser el número de iteraciones, la convergencia del algoritmo hacia el óptimo o bien cuando la población se encuentre estancada en un mínimo local.

modificaciones, extensiones y nuevas codificaciones al algoritmo genético básico con la finalidad de hacerlo más eficiente y adaptable a los problemas. Las modificaciones se centran principalmente en los operadores de selección, cruce y mutación, como se describirá enseguida.

1.2.1. Operadores de selección

La selección es un operador que merece gran cuidado ya que es el componente que determina el carácter del proceso de búsqueda evolutiva reflejando la *exploración* en el espacio de búsqueda o bien la *explotación* (o refinamiento) de los resultados obtenidos hasta la generación actual. En términos biológicos la diversidad genotípica de la población se relaciona con la distribución de la población sobre el espacio de búsqueda. A menor diversidad genotípica se obtiene una población mayormente concentrada sobre una región específica y la búsqueda se mantiene orientada en ruta, es decir, explotada. Por otro lado, una gran diversidad genotípica se relaciona con una población distribuida sobre todo el espacio, con lo cual la selección tiene una orientación volumétrica, o bien, explorativa. Las condiciones de la diversidad genotípica pueden ayudar a caracterizar la búsqueda en términos de la *presión de selección*, es decir, una “suave” presión de selección se ve reflejada en una mayor diversidad genotípica y una mayor exploración, lo cual conduce a una convergencia más confiable; por otro lado, una presión de selección “dura” se traduce en una menor diversidad genotípica, una mayor explotación y por lo tanto una mayor velocidad de convergencia.

En cada tipo de selección existen características que influyen en la presión de selección de los individuos, ya sea de forma volumétrica o de ruta. Ambas vertientes son indispensables en todo proceso evolutivo y el dominio de una de ellas en la búsqueda se traducirá en un mal desempeño del algoritmo genético.

Goldberg y Deb (1991), Bäck (1996) y Rudolph (1997) analizaron los operadores de selección empleando una medida conocida como “tiempo de asentamiento” (*takeover time*), la cual indica el número de iteraciones que necesita un algoritmo que sólo cuenta con selección para alcanzar una población cuyos individuos sean iguales. Esta idea está relacionada con la de presión de selección, de tal forma, que a un menor tiempo de asentamiento le corresponde una mayor presión de selección y a mayor tiempo de asentamiento le corresponde una menor presión de selección.

Existen muchos mecanismos de selección aplicados a los algoritmos evolutivos, algunos de los más importantes son la selección proporcional, la selección por rango (linear ranking), selección por torneo, la selección (μ, λ) y la selección $(\mu + \lambda)$.

La **selección proporcional** sin escalamiento es la que se aplica a los algoritmos genéticos básicos, su base es la probabilidad de seleccionar un individuo de la población mediante la técnica del conteo y se encuentra definida por la

expresión 1.5. Como se puede intuir, este tipo de selección no representa una gran presión de selección en los individuos y es por ello que se obtiene una convergencia más confiable. Sin embargo en [1] se puede encontrar la demostración de que para selecciones proporcionales con escalamiento se incrementa la probabilidad de elección de los individuos cuya aptitud se encuentre por arriba de la media y disminuye la probabilidad de elegir aquellos que se encuentren por debajo de la media, con lo que se aumenta la selección de mejores individuos.

El término de **rango lineal** o **linear ranking** se refiere a un método de selección el cual asigna una probabilidad que únicamente se guía por el rango del individuo sin importar cuál sea su función de aptitud. Entre las ventajas con las que cuenta esta selección se puede enunciar que puede acelerar la búsqueda considerablemente y que además la presión de selección se puede controlar con mayor facilidad que en el caso de la selección proporcional. El método original debe su nombre de rango lineal debido al uso de funciones lineales para mapear índices i de probabilidades de selección con sus respectivos valores esperados. En el caso de la presión de selección, ésta se puede variar mediante el ajuste del máximo valor esperado asignado a los individuos, el cual controla la pendiente de la función lineal.

En el caso de la **selección por torneo** de un grupo de q individuos seleccionados de forma aleatoria de la población, se elige el mejor de ellos como sobreviviente para la nueva generación. Este proceso es repetido tantas veces como sea necesario para llenar la siguiente población. Un valor común de q en el torneo es 2, en cuyo caso es conocido como *torneo binario*. Para torneos con valores de q mayores, es preciso indicar que la presión de selección aumenta conforme el valor de q crece. Por otro lado, comparada con el rango lineal, la selección por torneo siempre impone una presión de selección en el proceso de búsqueda, lo cual reduce considerablemente el tiempo de ejecución a tan sólo unas cuantas generaciones.

Las **selecciones** (μ, λ) y $(\mu + \lambda)$ difieren de las selecciones mostradas con anterioridad en tres importantes propiedades; primero, ambas son definidas bajo las suposiciones de que tanto los hijos como los padres son de tamaños diferentes (λ y μ , respectivamente), en segundo lugar, ambos métodos son completamente determinísticos, esto es, que no es necesario argumentar la selección mediante probabilidades, y finalmente, que ambos métodos excluyen a los peores individuos de las nuevas generaciones en lugar de mantenerlos con pequeña probabilidades. En el caso de la selección (μ, λ) , de una población de tamaño λ se seleccionan μ individuos, los cuales serán los padres de los λ hijos que conformarán la nueva población. En el caso de la selección (μ, λ) , la presión de selección se ve determinada por la razón λ/μ la cual aumenta o disminuye de forma proporcional con este factor. Para el caso de la selección $(\mu + \lambda)$, como se desea garantizar la supervivencia de los μ mejores padres de la generación anterior, se realiza una colecta de éstos y de los λ hijos generados, de tal forma que de la suma de ambos se reelijan a los λ mejores para la nueva generación.

Como es de esperarse, la presión de selección aumenta considerablemente con respecto a la selección (μ, λ) ya que la presencia de los mejores padres en la colecta disminuye por mucho el tiempo de ejecución del algoritmo.

1.2.2. Operadores de cruce

El operador de cruce es el elemento más importante en el Algoritmo Genético ya que permite preservar la exploración de los individuos en la región de búsqueda, impidiendo así la convergencia prematura hacia mínimos locales. Al igual que en el caso de los operadores de selección, existen muchos operadores de cruce, los cuales dependen de la representación de las cadenas que conforman a los individuos. Enfocándonos en el caso binario, como se vio en el algoritmo genético básico, existe la cruce de un punto y además, su generalización para dos o más puntos. En esta versión se eligen diversos puntos de cruce en los progenitores y se mezclan sus códigos genéticos de forma alternada para formar a los individuos de la nueva generación, obsérvese la figura 1.3. El caso más general es la conocida *cruza uniforme* en la cual cada uno de los bits de la cadena del nuevo individuo es tomado de uno de los dos padres, elegido de forma aleatoria.

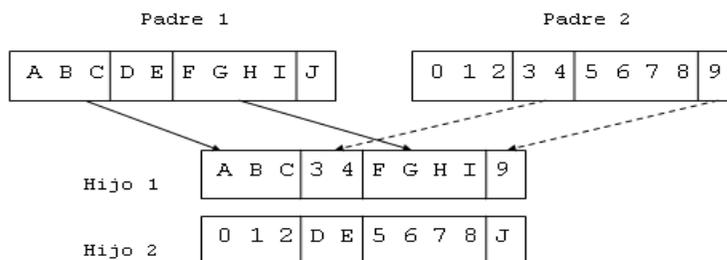


Figura 1.3: Cruza de cuatro puntos

Existen otros tipos de cruce que no únicamente toman dos padres para la recombinación, sino que emplean toda la población para conformar al nuevo individuo, como es el caso de la Cruza Simulada por Bits (**BSC**), la cual elige un individuo diferente para cada gen del hijo a formar.

En el caso continuo se puede encontrar una gran variedad de operadores de recombinación, en los cuales destacan aquellos que realizan *mezclas intermedias*. En éstas se toma una proporción α de los dos padres a recombinar para forman a los hijos, es decir:

$$h_{1_i} = p_{1_i} + \alpha(p_{2_i} - p_{1_i}) \quad (1.6)$$

$$h_{2_i} = p_{2_i} + \alpha(p_{1_i} - p_{2_i}) \quad (1.7)$$

En estas expresiones h_{1_i} y h_{2_i} son las variables i -ésimas de los dos hijos formados y p_{1_i} y p_{2_i} son las variables i -ésimas de los padres empleados para la recombinación. Un valor común de α es de 0.5, lo cual supone una aportación más o menos equivalente de los dos padres.

1.2.3. Operadores de mutación

El término de mutación resume una variedad de cambios (fuertes o débiles) en el código genético de los individuos. En general, la ventaja de las mutaciones tiene un efecto relativamente pequeño en el fenotipo. Y en efecto, debido a que las mutaciones son perturbaciones pequeñísimas en cada uno de los individuos, se podría pensar que una mutación es un cambio innecesario. Sin embargo, Holland (1975) hace una perfecta definición de este operador: “*La mutación es un operador secundario que asegura que el operador de cruce tenga un rango completo de modo que el plan adaptable no se quede atrapado en un mínimo local*”.

Bajo esta definición podemos entonces decir que el operador de mutación consigue mantener la diversidad en los individuos para que no exista un sólo gen dominante. Usualmente uno de cada mil genes es mutado y con ello es suficiente para imponerle la diversidad necesaria a la población, de hecho, perturbaciones mayores sólo generan destrucciones del código genético óptimo y conducen a la extinción de los buenos individuos.

De acuerdo con el tipo de cadena que se emplee para representar a los individuos podemos clasificar al operador de mutación en dos grandes grupos: discreto y continuo. En la representación discreta la mutación se realiza mediante cambios en cada uno de los genes de un individuo con cierta probabilidad p_M , de modo que un individuo mutado se observará como ejemplifica la figura 1.4.

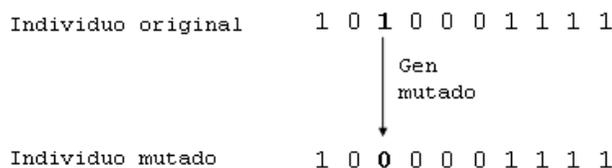


Figura 1.4: Mutación de un individuo con representación binaria

En el caso de la representación continua, el individuo mutado será una mezcla del individuo original y una perturbación ruidosa con cierta distribución, usualmente gaussiana. Supongamos que el individuo $\mathbf{x} = \{x_1, x_2, \dots, x_n\}$ será mutado con una pequeña perturbación normal con media cero y desviación estándar σ para dar lugar al nuevo individuo \mathbf{x}' , esto es

$$x'_i = x_i + N(0, \sigma_i), \quad i = 1, \dots, n \tag{1.8}$$

donde la perturbación se realiza sobre cada una de las variables del individuo y la desviación estándar de esa perturbación puede ser diferente para cada variable.

1.3. Bloques constructores y decepción

La teoría tradicional de los Algoritmos Genéticos asume que éstos funcionan por el descubrimiento, acentuación y recombinación de *bloques constructores*. Los bloques constructores son combinaciones de bits que confieren la mayor aptitud a las cadenas en las que están presentes.

Holland (1975) introduce la noción de *esquema* para formalizar el término de bloques constructores. Un esquema es un conjunto de bits que pueden ser descritos mediante una plantilla de unos, ceros y asteriscos, donde el asterisco simboliza un comodín. Por ejemplo, el esquema

$$H = 1 * * 1$$

encarna al conjunto de todas las cadenas de 4 bits que empiezan con uno y terminan con uno. Por notación, H (de hiperplano) representa un esquema y todas las cadenas que pueden ajustarse a estos esquemas se conocen como *instancias* del esquema, en este caso, por ejemplo, las instancias de H serán las cadenas 1001, 1011, 1101 y 1111.

El *teorema de los esquemas* fue propuesto por Holland (1975) y explica que los bloques cortos y de bajo orden cuya aptitud promedio se mantiene por encima de la media recibirán un número de muestras que crecen exponencialmente con el tiempo [3]. Sin embargo, pese a que el teorema de los esquemas lidia con la cruce y la mutación por ser altamente disruptivas, lo que hace funcionar al Algoritmo Genético, es precisamente su capacidad de recombinar las instancias de los esquemas favorables para formar otras igualmente buenas o incluso mejores. Esta suposición se conoce como la *hipótesis de los bloques constructores* y fue propuesta por Goldberg (1989).

Ahora bien, la hipótesis de los bloques constructores dice que la cruce combina esquemas cortos y de alto desempeño para formar candidatos más aptos, pero no da una descripción de cómo ocurre esta combinación. Para investigar más detalladamente el comportamiento de los esquemas y su recombinación se diseñaron ciertas funciones llamadas *de la carretera real* las cuales intentan capturar la esencia de los bloques constructores en una forma idealizada (ver [4]). Estas funciones estaban construidas de tal forma que cumplieran la hipótesis de los bloques constructores (cadenas cortas, de bajo orden y altamente aptos) de modo que el algoritmo genético llegaría fácilmente a la solución. Sin embargo, se observó que esto realmente no era cierto y que incluso algoritmos que se esperaba que tuvieran dificultad de llegar a la solución, la alcanzaron hasta 10 veces más rápido que el Algoritmo Genético.

Estos resultados mostraron la necesidad de proponer un nuevo Algoritmo Genético idealizado que evitara problemas con respecto a la hipótesis de los bloques constructores, éste tendría que poseer las siguientes características:

1. La población tiene que ser lo suficientemente grande, el proceso de selección debe ser lo suficientemente lento y el porcentaje de mutación debe ser suficientemente alto para asegurar que no existan ninguna posición que permanezca fija con un solo valor en ninguna cadena.
2. La selección debe ser lo suficientemente fuerte como preservar los esquemas deseables encontrados.
3. El porcentaje de cruce debe ser tal que el tiempo en que ocurra una cruce que combine dos esquemas deseados sea pequeño con respecto al tiempo de descubrimiento de los esquemas deseados.

sin embargo, como se puede apreciar, existen mecanismos que no son compatibles entre sí y que hablan sólo de un Algoritmo ideal y no práctico.

1.3.1. Los problemas de los AG: las funciones deceptivas

La idea básica tras el diseño de funciones deceptivas para un algoritmo genético es violar de manera extrema la hipótesis de los bloques constructores. En otras palabras, se busca que los bloques cortos y de bajo orden conduzcan a bloques constructores largos y de mayor orden que sean subóptimos.

Un ejemplo de función deceptiva se conoce como “función tramposa separable de orden 5” (trap_5). En este problema se asume que el tamaño de la cadena es un múltiplo de cinco y que se encuentra dividida en partes disjuntas de 5 bits cada una. Cada parte se ajusta en cada corrida del algoritmo y éste desconoce el comportamiento de cada partición en la cadena. Denótese las posiciones correspondientes a la i -ésima partición mediante los índices $b_{i,1}$ a $b_{i,5}$. Y defínase

$$f_{\text{trap}_5}(x) = \sum_{i=1}^{\frac{n}{5}} \text{trap}_5(x_{b_{i,1}} + \dots + x_{b_{i,5}}) \quad (1.9)$$

donde

$$\text{trap}_5(u) = \begin{cases} 5 & \text{si } u = 5 \\ 4 - u & \text{otro caso} \end{cases} \quad (1.10)$$

En esta expresión u denota el número de unos en el bloque de 5 bits. Como se puede apreciar, este problema tiene un óptimo global en $u = 5$ y un óptimo local en $u = 0$. Cuando se trabajan con cadenas de n bits sólo se cuenta con un óptimo global en una cadena con n unos y $2^{\frac{n}{5}} - 1$ óptimos locales en al menos una partición de la cadena que se encuentre en puros ceros. Refiérase a [2] para un estudio más profundo del comportamiento de este problema. En este caso, la función trap_5 es completamente *deceptiva*, es decir, que la aptitud de cualquier bloque de bits de orden menor que 5 la aleja del óptimo global.

Pese a que la función trap_5 es completamente artificial, sugiere la existencia de problemas que el Algoritmo Genético no puede resolver, y esto debido principalmente a la destrucción de los bloques que construyen al óptimo (*bloques constructores*), mediante técnicas de cruce o mutación. Si la cruce empleada combinara eficientemente las soluciones promisorias de cada cadena, habría mayor probabilidad de mantener los bloques buenos y de esta manera alcanzar el óptimo.

A raíz de esta clase de problemas, los cuales carecen de solución en el Algoritmo Genético, se han propuesto diversas formas para atacar a los problemas deceptivos. Entre estas formas se incluyen métodos que aprenden la estructura del problema y emplean esta información para asegurar una mezcla más apropiada y así aumentar los bloques constructores. Algunas de estas técnicas están basadas en modelos probabilísticos que pretenden guiar la búsqueda hacia zonas más prometedoras en lugar de usar la cruce y mutación como en el Algoritmo Genético Básico.

Una técnica a considerar es trabajar las probabilidades de aparición de las variables para regenerar a los nuevos individuos en lugar de trabajar en sí con los individuos de la población. Esta técnica es conocida como Algoritmos de Estimación de Distribución y se tratará con mayor profundidad en el capítulo siguiente.

1.4. Conclusiones del capítulo 1

En el presente capítulo se dieron las bases de los Algoritmos Evolutivos, los cuales surgen como una imitación al comportamiento evolutivo de las especies empleando ciclos y algoritmos computacionales para tal fin; se establece la representación de los individuos mediante cadenas discretas o continuas y la forma básica de implementar un Algoritmo Genético. Además, se estudiaron los diversos operadores de selección, cruce y mutación para el caso discreto y continuo, así como la eficiencia de los operadores de selección en términos del tiempo de asentamiento. Se introdujo la definición de *esquema e instancia* y se enunciaron tanto el teorema de los esquemas como la hipótesis de los bloques constructores. Se dió un ejemplo de las funciones deceptivas, las cuales representan una gran dificultad para los Algoritmos Genéticos debido a lo disruptivo de la cruce y a la pérdida de los bloques que forman la solución. Finalmente se introduce un nuevo tipo de Algoritmo Evolutivo, el cual emplea la distribución de los buenos individuos para reorientar a la población y regenerarla basándose en estas estadísticas y no en cruces y mutaciones como lo hace el Algoritmo Genético.

Capítulo 2

Algoritmos con Estimación de Distribución

El comportamiento de los Algoritmos Evolutivos vistos en el primer capítulo depende en gran medida del ajuste de un gran conjunto de parámetros (el tamaño de la población, el factor de cruce, mutación, número de generaciones, etc.) que ya representan en sí un problema de optimización. Esta razón, junto con el deseo de predecir el comportamiento de los individuos en la población motivaron el estudio de un nuevo tipo de algoritmos que emplearan información estadística sobre la población para realizar las predicciones. Estos algoritmos son conocidos como Algoritmos de Estimación de Distribución (EDAs) y fueron introducidos por Mühlenbein, Paaß y Baluja (1996).

En EDAs no existe ni cruce ni mutación como en el caso de los Algoritmos Genéticos, en su lugar la nueva población se remuestrea a partir de la distribución de una muestra. Por otro lado, en Algoritmos Genéticos, las asociaciones entre las variables eran tomadas a partir de los bloques constructores y estaban implícitas en la representación de los individuos, mientras que en EDAs las asociaciones se encuentran explícitamente expresadas mediante la función de distribución conjunta de los individuos seleccionados.

En este segundo capítulo se revisarán los conceptos básicos de probabilidad que sirven como base de los Algoritmos de Estimación de Distribución, la terminología aplicada a lo largo de la tesis, definiciones de densidad y distribución, así como sus características principales. Se enunciarán los algoritmos conocidos que cuentan con una base paramétrica tanto en su representación real como discreta y sus principales características, además se estudiarán la bases de los algoritmos no paramétricos y los primeros intentos para su elaboración. Finalmente se concluye con algunas revisiones y observaciones del capítulo.

2.1. Conceptos de probabilidad

Debido a la naturaleza de los Algoritmos de Estimación de Distribución es preciso asentar las bases probabilísticas que sustentan el desarrollo de estos algoritmos. En primera instancia, defínase una *variable aleatoria* como una función real de los elementos de un espacio muestral S . Las variables aleatorias se representan mediante letras mayúsculas (por ejemplo X, Y o Z) y cualquier valor que tome esta variable aleatoria se escribirá en minúsculas (x, y o z). Considérense las siguientes definiciones:

Definición 2.1 (Probabilidad marginal). *La probabilidad marginal sobre la i -ésima variable se obtiene mediante*

$$p(x_i) = \sum_{x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_n} p(x_1, \dots, x_n)$$

En este caso $p(x_1, \dots, x_n)$ es la *probabilidad conjunta* o bien la probabilidad de que las variables aleatorias X_1, \dots, X_n tomen los valores x_1, \dots, x_n , respectivamente. A esta correspondencia se le conoce como *evento* y representa un subconjunto del espacio muestral. La probabilidad de que ocurra un evento X dado otro evento Y se indica en la definición 2.2.

Definición 2.2 (Probabilidad condicional). *Dadas las variables X e Y , la probabilidad condicional para los valores x de X e y de Y , se define como*

$$p(x|y) = \frac{p(x, y)}{p(y)}, \quad p(y) \neq 0 \quad (2.1)$$

Definición 2.3 (Independencia condicional). *Se dice que la variable aleatoria X es marginalmente independiente de la variable aleatoria Y , si y sólo si para todos los valores x de X e y de Y se satisface que*

$$p(x|y) = p(x), \quad p(y) \neq 0 \quad (2.2)$$

Siempre que se cumpla la definición 2.3 también se cumplirá que $p(x, y) = p(x)p(y)$, y esto puede verificarse sustituyendo la ecuación 2.2 en la expresión 2.1.

Definición 2.4 (Función de distribución acumulada). *Sea X una variable aleatoria y $P\{X \leq x\}$ la probabilidad de que esta variable aleatoria tome valores menores a x , defínase entonces*

$$F_X(x) = P\{X \leq x\} \quad (2.3)$$

como la *función de distribución acumulada de la variable aleatoria X* .

La función de distribución acumulada tiene ciertas propiedades las cuales se derivan del hecho de que $F_X(x)$ es una probabilidad, algunas de estas propiedades se enuncian enseguida:

$$F_X(-\infty) = 0 \quad (2.4a)$$

$$F_X(\infty) = 1 \quad (2.4b)$$

$$F_X(x_1) \leq F_X(x_2) \quad \text{si} \quad x_1 < x_2 \quad (2.4c)$$

$$F_X(x^+) = F_X(x) \quad (2.4d)$$

La propiedad 2.4d se refiere a que la función de distribución acumulada debe ser continua por la derecha y la 2.4c que debe ser una función creciente. Todas estas propiedades deben emplearse para probar si una función cualquiera $G_X(x)$ es una función de distribución válida. Existen diversos tipos de distribuciones, pero la más común es la distribución normal ya que todo conjunto grande de datos puede ser modelado con esta distribución y además facilita los cálculos.

Definición 2.5 (Función de densidad). *Sea X una variable aleatoria y $F_X(x)$ su función de distribución acumulada, defínase entonces*

$$f_X(x) = \frac{dF_X(x)}{dx} \quad (2.5)$$

como la función de densidad de la variable aleatoria X .

Dos propiedades interesantes de la función de densidad se enuncian enseguida:

$$f_X(x) \geq 0 \quad \text{para toda } x \quad (2.6a)$$

$$\int_{-\infty}^{\infty} f_X(x) dx = 1 \quad (2.6b)$$

Ambas propiedades deben ser usadas para probar si una función cualquiera $g_X(x)$ es una función de densidad válida. Tanto la función de densidad como la función de distribución pueden emplearse para obtener cualquier información sobre la variable aleatoria en cuestión, en el caso particular de los EDAs se emplea la función de distribución para modelar a los individuos.

Las definiciones 2.4 y 2.5 pueden ser extendidas para el caso general de n variables aleatorias. En este caso las funciones se conocen como *función de distribución acumulada conjunta* y *función de densidad conjunta*, respectivamente. En la definición 2.6 se extiende la definición de la función de distribución acumulada para el caso de n variables.

Definición 2.6 (Función de distribución acumulada conjunta). *Sea X_1, \dots, X_n una vector de n variables aleatorias y $P\{X_1 \leq x_1, \dots, X_n \leq x_n\}$ la probabilidad de que cada elemento del vector de variables aleatorias tome valores menores a x_1, \dots, x_n , respectivamente. Defínase entonces*

$$F_{X_1, \dots, X_n}(x_1, \dots, x_n) = P\{X_1 \leq x_1, \dots, X_n \leq x_n\} \quad (2.7)$$

como la función de distribución acumulada conjunta del vector de variables aleatorias X_1, \dots, X_N .

El término de *esperanza* se aplica cuando en el proceso de promediar se involucra a una variable aleatoria. Si hablamos de la variable aleatoria X , se emplea la notación $E[X]$ o bien \bar{X} para referirse al valor esperado de la variable aleatoria X . Refiérase a la siguiente definición:

Definición 2.7 (Valor esperado). *Sea X una variable aleatoria y $f_X(x)$ su correspondiente función de densidad, defínase entonces*

$$E[X] = \bar{X} = \int_{-\infty}^{\infty} x f_X(x) dx$$

como el valor esperado de la variable aleatoria X .

Una aplicación inmediata del valor esperado es el cálculo de momentos. Existen dos clases de momentos; los momentos con respecto al origen y los momentos centrales. Para este trabajo de investigación son de interés los momentos centrales, los cuales se definen a continuación:

Definición 2.8 (Momento central). *Sea X una variable aleatoria y \bar{X} la esperanza de la variable aleatoria X , defínase entonces*

$$\mu_n = E[(X - \bar{X})^n]$$

como el n -ésimo momento central de la variable aleatoria X .

El segundo momento central es tan importante que se le da la notación σ_X^2 y se conoce como la *varianza* de X , además, la raíz positiva de este valor σ_X se conoce como la *desviación estándar* de X y es una medida de la dispersión de los datos con respecto a la media, en la función de densidad.

Al trabajar con dos variables aleatorias X e Y ya se puede hablar de *momentos centrales conjuntos*, los cuales se encuentran definidos como sigue:

Definición 2.9 (Momento central conjunto). *Sean X e Y dos variables aleatorias y \bar{X} e \bar{Y} sus correspondientes valores esperados, defínase entonces*

$$\mu_{nk} = E[(X - \bar{X})^n (Y - \bar{Y})^k]$$

como el momento conjunto de orden $n + k$ de las variables aleatorias X e Y .

El momento conjunto de segundo orden μ_{11} es muy importante y es conocido como la *covarianza* de X e Y y está dado por el símbolo Σ_{XY} . Éste denota las relaciones entre las dos variables, por ejemplo, si X e Y fueran independientes o decorrelacionadas entonces su covarianza sería cero.

Definición 2.10 (Coeficiente de correlación). *Sean X e Y dos variables aleatorias y σ_X y σ_Y sus correspondientes desviaciones estándar. Además, sea Σ_{XY} la covarianza de las variables X e Y , defínase entonces*

$$\rho = \frac{\Sigma_{XY}}{\sigma_X \sigma_Y} \quad (2.8)$$

como el coeficiente de correlación de las variables X e Y .

Al igual que la covarianza, el coeficiente de correlación indica las relaciones entre las variables involucradas, pero debido a que se encuentra normalizado ($-1 \leq \rho \leq 1$), se emplea con frecuencia para cuantificar la fuerza de las relaciones entre las variables. El valor absoluto de ρ indica qué tan relacionada de encuentra una variable con respecto a la otra y el signo indica el tipo de relación.

2.2. Algoritmos de Estimación de Distribución

Los Algoritmos de Estimación de Distribución son un tipo especial de algoritmos evolutivos que emplean probabilidades y estadísticos para la regeneración de la población. Obsérvese el algoritmo 2 el cual muestra el pseudocódigo de un Algoritmo de Estimación de Distribución en su forma básica.

Algoritmo 2 EDA

- 1: $t=0$
 - 2: Generar N individuos de forma aleatoria y guardarlos en $P(t)$.
 - 3: **while** no se cumpla la condición de paro **do**
 - 4: Evaluar cada uno de los individuos de $P(t)$.
 - 5: Seleccionar $M \leq N$ individuos de $P(t)$ y colocarlos en la muestra S .
 - 6: Generar la nueva población $P(t+1)$ a partir de la distribución de probabilidad de S .
 - 7: $t=t+1$
 - 8: **end while**
-

Al observar el algoritmo 2 se puede apreciar el parecido con un Algoritmo Genético convencional, sin embargo, difiere de éste en el punto 6, es decir, en la generación de la población a partir la distribución de probabilidad. Existen diversos tipos de EDAs y cada uno de ellos emplea una técnica diferente para realizar la aproximación de la distribución conjunta.

Existen modelos que suponen independencia entre las variables, con lo cual la distribución conjunta queda reducida al producto de las probabilidades de cada variable; otros modelos suponen que los datos provienen de una distribución normal y emplean fórmulas ya establecidas para el cálculo de los estadísticos necesarios para la regeneración de los datos. En otros tantos se suponen dependencias de segundo o tercer orden e incluso modelos más complejos y otros pocos asumen que los datos no provienen de una distribución conocida y regeneran la población a partir de los datos mismos.

Debido a la gran variedad de Algoritmos de Estimación de Distribución, se suele realizar la clasificación de ellos en dos grandes grupos para facilitar su estudio: los algoritmos con representación discreta y los algoritmos con representación real. En ambos casos existen EDAs sin dependencias o bien con dependencias que pueden ser bivariadas o múltiples, en el caso discreto; y paramétricos o no paramétricos, en el caso continuo. En la figura 2.1 se presenta un diagrama de clasificación de los EDAs de acuerdo con esta estructura. La parte baja del

organigrama muestra ejemplos de los EDAs representantes de cada categoría. Acentuado con otro color se indica el área de investigación de interés plasmado en este trabajo de tesis. Note que no se muestra un algoritmo representativo en esta sección, y esto debido al escaso trabajo de investigación realizado hasta la actualidad, sin embargo, más adelante, se hablará de los primeros intentos por explorar esta área y se enunciarán algunos ejemplos.

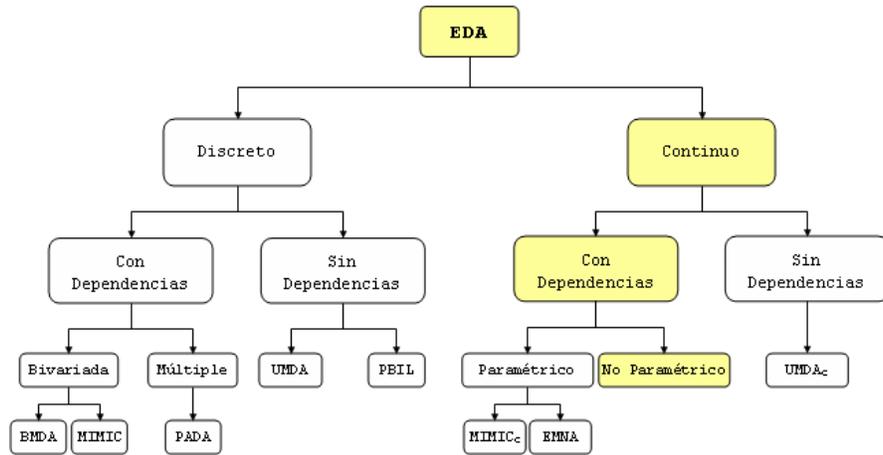


Figura 2.1: Estructura de clasificación de los EDAs

2.2.1. EDAs con representación discreta

Los Algoritmos con estimación de distribución pertenecientes a esta clasificación pueden estar representados mediante modelos gráficos, como en el caso de los algoritmos con dependencias, o bien, ser tan simples que supongan independencia entre las variables y se reduzcan sólo al producto de las probabilidades marginales. A continuación se analizarán los modelos discretos sin dependencias y aquellos con dependencias bivariadas y múltiples.

Modelos independientes

En esta sección se introducen dos ejemplos de modelos que consideran que la función de distribución acumulada conjunta es el producto de las n probabilidades marginales de las variables que componen al vector n -dimensional de variables aleatorias. Aunque esta suposición reduce significativamente los cálculos, no puede existir nada más alejado de la realidad ya que usualmente existen dependencias entre las variables.

El Algoritmo de Distribución Marginal Univariada **UMDA** fue introducido por Mühlenbein en 1998. En este algoritmo la distribución de probabilidad

conjunta $p_{X_1, \dots, X_n}^t(x_1, \dots, x_n)$ está factorizada como el producto de las distribuciones marginales univariadas y se describe por la siguiente expresión:

$$p_{X_1, \dots, X_n}^t(x_1, \dots, x_n) = \prod_{i=1}^n F_{X_i}^t(x_i)$$

Cada distribución marginal univariada es estimada a partir de las frecuencias marginales:

$$p_{X_i}^t(x_i) = \frac{\sum_{j=1}^M \delta_j(X_i = x_i | S_{t-1})}{M}$$

donde

$$\delta_j(X_i = x_i | S_{t-1}) = \begin{cases} 1 & \text{si , } X_i = x_i \text{ en el } j\text{-ésimo bit de } D_{t-1} \\ 0 & \text{otro caso.} \end{cases}$$

y D_{t-1} es el conjunto de muestra en la iteración $t - 1$.

El Algoritmo de Aprendizaje Incremental Basado en la Población **PBIL** fue introducido por Baluja en 1994. El algoritmo emplea para su funcionamiento un vector de probabilidades de tamaño n cuyos valores corresponden a la probabilidad de obtener un 1 en la i -ésima variable de la población P , esto es:

$$p^t(\mathbf{x}) = p^t(x_1) \cdots p^t(x_i) \cdots p^t(x_n)$$

Este vector de probabilidades se emplea para generar a los individuos en cada nueva iteración t y los M individuos mejores son seleccionados para actualizar el vector de probabilidades haciendo uso de una combinación convexa:

$$p^{t+1}(\mathbf{x}) = (1 - \alpha)p^t(\mathbf{x}) + \alpha \frac{1}{M} \sum_{k=1}^M \mathbf{x}_{k:N}^t$$

donde $\mathbf{x}_{k:N}^t$ denota el k -ésimo elemento de la muestra S tomado del conjunto de N individuos de la población y $\alpha \in (0, 1]$ es un parámetro del algoritmo (vea [5]). Como se puede verificar, el algoritmo del PBIL coincide con el UMDA en el caso especial de que $\alpha = 1$.

Dependencias bivariadas

En el caso de los EDAs con dependencias bivariadas se realiza una aproximación de la distribución conjunta tomando en cuenta dependencias entre cada dos variables. Tanto en los métodos que consideran dependencias bivariadas como aquellos que las consideran multivariadas y tienen representación discreta, se emplean nodos que simbolizan a cada variable y arcos que indican la relación entre éstas. En conjunto se trabaja con grafos los cuales aumentan el grado de dificultad de los cálculos conforme se suponen dependencias de mayor orden.

El algoritmo de Maximización de Información Mutua mediante Clústeres de Entrada (**MIMIC**) realiza en cada generación una búsqueda de la mejor permutación entre las variables con la finalidad de obtener la función de distribución acumulada conjunta $p_{X_1, \dots, X_n}^{\pi}(x_1, \dots, x_n)$ que más se aproxime a la distribución empírica del conjunto de puntos seleccionados. Para cotejar qué tan buena es una permutación, se emplea la discrepancia de *Kullback-Leibler* como medida de la semejanza entre dos determinadas distribuciones.

Definición 2.11 (Discrepancia de Kullback-Leibler). *Sea el vector de variables aleatorias $\mathbf{X} = (X_1, \dots, X_n)$ y sean $P(\mathbf{X})$ y $P_a(\mathbf{X})$ dos distribuciones de probabilidad del vector de variables aleatorias \mathbf{X} . Defínase entonces*

$$D(P \parallel P_a) = \sum_{\mathbf{X}} P(\mathbf{X}) \log \frac{P(\mathbf{X})}{P_a(\mathbf{X})} \quad (2.9)$$

como la discrepancia de *Kullback-Leibler* entre las distribuciones P y P_a .

Es posible ver que esta discrepancia tiene la propiedad de que $D(P \parallel P_a) \geq 0$, y sólo se cumple la igualdad siempre que ambas distribuciones sean iguales.

Aproximamos entonces:

$$p_{\mathbf{X}}^{\pi}(\mathbf{x}) = p(x_{i_1}|x_{i_2}) \cdot p(x_{i_2}|x_{i_3}) \cdots p(x_{i_{n-1}}|x_{i_n}) \cdot p(x_{i_n})$$

y $\pi = (i_1, i_2, \dots, i_n)$ denota la permutación de los índices $1, 2, \dots, n$. Con lo cual la búsqueda del algoritmo se reduce a la búsqueda de la mejor permutación que minimice 2.9. De Bone (1997) propone una técnica para reducir la búsqueda de esta permutación mediante la elección de aquellas variables que minimicen su entropía estimada dada una anterior. Para la elección de la primera variable, se elige la que cuente con menos entropía marginal y posteriormente se le adhieren aquellas que mantengan mínima la entropía condicional dada la variable inmediata anterior.

El Algoritmo de Distribución Marginal Bivariada (**BMDA**) propuesto por Pelikan y Mühlenbein (1999) propone una factorización de la función de distribución conjunta que sólo requiere de estadísticos de segundo orden. El primer paso del algoritmo es la selección de una variable aleatoria la cual sirva como base de una gráfica acíclica no necesariamente conectada conocida como *gráfica de dependencias*. Esta variable se elige como aquella que posea la máxima dependencia marginal de acuerdo con la medida χ^2 . Posteriormente se van conectando las variables del grupo restante que tengan la mayor dependencia con aquellas puestas previamente y se repite este proceso hasta que ya no existan variables del grupo restante que tengan una dependencia mayor a cierto límite que les permita unirse al grafo. Si aún existen variables que no se han colocado en la gráfica, entonces se repite el proceso desde el inicio y termina cuando todas las variables están incluidas en la gráfica de dependencias. La prueba χ^2 se define enseguida:

Definición 2.12 (Prueba χ^2). Sean X e Y dos variables aleatorias y N el número de datos que se tienen de cada variable aleatoria. Sea además $p_X(x)$ la probabilidad marginal de la variable aleatoria X y $p_{X,Y}(x,y)$ la probabilidad conjunta entre las dos variables X e Y . Defínase entonces:

$$\chi_{X,Y}^2 = \sum_{x,y} \frac{(Np_{X,Y}(x,y) - Np_X(x)p_Y(y))^2}{Np_X(x)p_Y(y)}$$

como la prueba de estadística de independencia χ^2 entre las variables X e Y .

Si en la definición 2.12 las variables X e Y son independientes con probabilidad del 95 %, entonces la prueba chi-cuadrada satisface la siguiente desigualdad:

$$\chi_{0.05}^2 < 3.84$$

con un grado de libertad ($df=1$). Este valor se emplea como límite para averiguar si las dos variables testeadas son lo suficientemente dependientes una de otra como para ser parte del grafo de dependencias. De no haber otra variable que supere este valor, se busca una nueva base (o nodo raíz) para formar el siguiente árbol. El grafo final tendrá una apariencia similar a la mostrada en la figura 2.2.

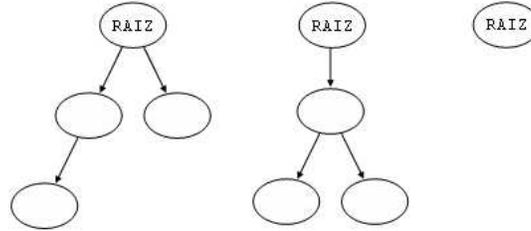


Figura 2.2: Representación gráfica del modelo de probabilidad del BMDA

Como se advierte en esta imagen la apariencia de este algoritmo es la de un conjunto de árboles con la mayor dependencia entre cada ramificación, en contraste, el MIMIC se observará como una única cadena de variables enlazadas, en cada iteración t .

Además de estos modelos se puede citar el **árbol de dependencias** propuesto por Chow y Liu en 1968, el cual es la base del BMDA. En este modelo se utiliza la distancia de Kullback-Leibler como una medida de semejanza entre dos distribuciones, denotando como un *árbol de máximo peso* a aquel que posea la máxima información mutua entre las variables del grafo. De esta manera, el árbol con máxima información mutua será también el que minimice la distancia de Kullback y por ende aproxime mejor la distribución empírica, vea [6].

Dependencias múltiples

Los EDAs que emplean más de dos dependencias para aproximar la función de distribución conjunta son variados y uno de los primeros trabajos realizados al respecto fue el propuesto por Baluja y Davies (1998), desafortunadamente, en este trabajo sólo se describe la posibilidad de su funcionamiento pero no se llega a implementarlo.

En estos modelos, al igual que en el caso bivariado, se trabajan con representaciones gráficas para indicar la asociación de las variables y con arcos que pueden ser unidireccionales o bidireccionales de acuerdo con el tipo de EDA que se implemente.

El Algoritmo de Aproximación de Distribución por Poliárboles **PADA** emplea una red Bayesiana para factorizar la función de distribución conjunta. El algoritmo consiste en generar *poliárboles* (es decir, estructuras que no contienen más de un arco unidireccionado por cada dos variables) a partir de dependencias o independencias entre las variables. El criterio de dependencia en este modelo es la *d-separabilidad* la cual puede considerarse como la definición gráfica de la independencia condicional [7]. Observe en la figura 2.3 la representación gráfica del modelo de probabilidad del PADA.

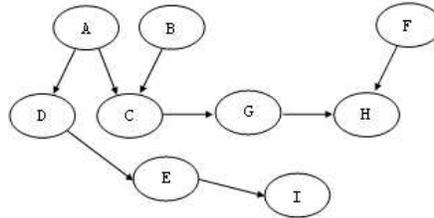


Figura 2.3: Representación gráfica del modelo de probabilidad del PADA

Además de este modelo, existen representaciones más complejas que permiten ciclos no dirigidos, las cuales son conocidas como *redes Bayesianas multiconectadas* y dependiendo de si son simples o generales pueden admitir tanto un tipo especial de ciclo no dirigido (donde cada par de nodo con un hijo común no tiene antecesores comunes, ni uno es antecesor de otro), como todo tipo de ciclo no dirigido.

2.2.2. EDAs con representación real

La gran mayoría de los EDAs con representación real es una simple extensión de algunos de los algoritmos con representación discreta como lo son el UMDA, el PBIL, el MIMIC, etc., y otros son nuevas propuestas que, en su gran mayoría, emplean funciones gaussianas para aproximar la distribución conjunta. De cualquier forma, en la presente sección se le echa un vistazo a algunos de los algoritmos más representativos en el ámbito real con y sin dependencias y se

introduce un nuevo modelo para aproximar la distribución conjunta de forma no paramétrica que se empleará a lo largo de este trabajo de investigación.

Modelos independientes

En este tipo de modelos es común suponer que no existe dependencia entre las variables y que la función de densidad conjunta es el resultado del producto de las densidades independientes normales de cada variable. Con lo cual se denota $\mathbf{X} \sim \mathcal{N}(x; \mu, \sigma)$, es decir, que el vector de variables aleatorias \mathbf{X} se *distribuye* de forma normal con parámetros μ y σ .

El Algoritmo de Distribución Marginal Univariada para el dominio continuo **UMDA_c** realiza algunas pruebas estadísticas para encontrar la función de densidad que mejor se ajuste a la densidad variable. En el caso de que la densidad se modele con una función normal, entonces se deben calcular los dos parámetros μ y σ a partir de los estimadores de máxima verosimilitud, esto es:

$$\hat{\mu}_i = \bar{X}_i, \quad \hat{\sigma}_i = \sqrt{\frac{1}{N} \sum_{r=1}^N (x_{i,r} - \bar{X}_i)^2} \quad (2.10)$$

En el Algoritmo de Aprendizaje Incremental Basado en la Población para el espacio continuo **PBIL_c** se asume que la función de densidad conjunta de la población sigue una distribución gaussiana y que es factorizable como el producto de las densidades marginales independientes. Cada uno de los elementos del vector de medias es actualizado mediante la ecuación 2.11

$$\mu_i^{t+1} = (1 - \alpha)\mu_i^t + \alpha[x_{ibest,1}(t) + x_{ibest,2}(t) - x_{iworst}(t)] \quad (2.11)$$

donde μ_i^{t+1} representa el i -ésimo componente del vector de medias en la generación $t + 1$ y $x_{ibest,1}(t)$, $x_{ibest,2}(t)$, $x_{iworst}(t)$ representan el mejor, el segundo mejor y el peor individuo de la generación t , respectivamente. Al igual que en el caso discreto α es una constante. Para adaptar el vector de varianzas se proponen cuatro heurísticas: (1) usar un valor constante, (2) ajustarlas como en la estrategia (1, λ), (3) calcular la varianza muestral de los mejores k individuos o bien (4) empleando una combinación convexa al igual que en el caso de las medias.

Modelos paramétricos con dependencias

Los modelos con representación real y dependencias, al igual que en el caso discreto, muestran dependencias bivariadas o múltiples pero en todos ellos se supone que los datos provienen de una distribución normal y es con ello con lo que se hace la estimación de los parámetros μ y σ . Es precisamente por esto que se ha decidido clasificar a los modelos con representación real y dependencias como paramétricos y no paramétricos.

En el caso de los modelos paramétricos con dependencias bivariadas podemos encontrar el **MIMIC**_C^G en el cual se asume que la distribución de probabilidad para cada par de variables es una Gaussiana bivariada. Como en el caso discreto, en el MIMIC continuo se emplean entropías para minimizar la distancia de Kullback-Leibler. Recurriendo al teorema de Whittaker (1990), el cual utiliza un vector de variables \mathbf{X} que se supone se distribuye de forma normal con media μ y matriz de covarianzas Σ y aplicándolo a la función de distribución normal bivariada, se encuentra que la entropía de \mathbf{X} está dada por:

$$h(\mathbf{X}) = \frac{1}{2}(1 + \log 2\pi) + \log \sigma_X$$

y que la entropía condicional se calcula como:

$$h(X|Y) = \frac{1}{2} \left[(1 + \log 2\pi) + \log \left(\frac{\sigma_X^2 \sigma_Y^2 - \sigma_{XY}^2}{\sigma_Y^2} \right) \right]$$

Con lo cual el algoritmo del MIMIC continuo funciona haciendo uso de dos pasos principales. En el primero de ellos se elige la variable con la menor varianza muestrada (por ejemplo Y), y en el segundo paso se elige aquella otra variable (por ejemplo, X) cuya estimación de $\frac{\sigma_X^2 \sigma_Y^2 - \sigma_{XY}^2}{\sigma_Y^2}$ con relación a la variable anterior Y , sea menor. Finalmente esta última variable se enlaza con la primera y el proceso es repetido de forma condicional hasta terminar todas las variables.

Cuando el número de dependencias aumenta, el número de parámetros a calcular crece de forma proporcional con la dimensión del problema, sin embargo, modelos como el **EMNA**_{global} realizan esta búsqueda con cálculos muy simples y de bajo costo computacional. En este algoritmo se realiza una aproximación de la función de distribución normal multivariada en cada generación mediante la estimación del vector de medias $\boldsymbol{\mu}^t = (\mu_1^t, \dots, \mu_n^t)$ y la matriz de covarianzas $\boldsymbol{\Sigma}$. Estos parámetros emplean los estimadores de máxima verosimilitud para su aproximación:

$$\begin{aligned} \hat{\mu}_i &= \bar{X}_i & i &= 1, \dots, n \\ \hat{\sigma}_i^2 &= \frac{1}{N} \sum_{r=1}^N (x_{i,r} - \bar{X}_i)^2 & i &= 1, \dots, n \\ \hat{\sigma}_{i,j}^2 &= \frac{1}{N} \sum_{r=1}^N (x_{i,r} - \bar{X}_i)(x_{j,r} - \bar{X}_j) & i, j &= 1, \dots, n \quad i \neq j \end{aligned}$$

Posteriormente hace uso de estos parámetros para regenerar la población suponiéndola, asimismo, normal con parámetros $\hat{\boldsymbol{\mu}}$ y $\hat{\boldsymbol{\Sigma}}$. Para más detalle sobre estos algoritmos vea [8].

Modelos no paramétricos

Pese a que en las secciones previas ya se ha hablado de los algoritmos no paramétricos con dependencias bivariadas y múltiples, en el caso de los algoritmos no paramétricos aún no se ha llegado a un modelo que tome en cuenta las dependencias entre las variables, tal el caso de un algoritmo evolutivo que emplea histogramas marginales para la aproximación de la función de densidad, propuesto por Tsutsui et al. en 2001. En este trabajo se consideran dos tipos de histogramas: los histogramas ajustados por ancho (FWH) y los ajustados por alto (FHH).

En FWH se divide el espacio de búsqueda $[\text{mín}_j, \text{máx}_j]$ de cada variable $x_j (j = 1, \dots, n)$ en $H (h_j = 0, 1, \dots, H - 1)$ secciones. Con lo cual la función de densidad estimada $f_{FWH}^j[h]$ de cada sección h_j es:

$$f_{FWH}^j[h] = \frac{\left| \left\{ V[i][j] \mid i \in \{0, \dots, N - 1\} \wedge h \leq \frac{V[i][j] - \text{mín}_j}{\text{máx}_j - \text{mín}_j} H < h + 1 \right\} \right|}{N} \quad (2.12)$$

donde $V[i][j]$ es el valor de la variable x_j del individuo i en la población de tamaño N . Para generar a la nueva población, primero se selecciona una sección empleando la probabilidad expresada por la ecuación 2.12 y luego se genera el individuo de forma aleatoria uniforme dentro de los límites de la sección escogida.

Si bien en FWH las secciones tiene el mismo ancho, en FHH las secciones tienen la misma altura, lo cual significa que cada sección contiene el mismo número de puntos. Debido a que la probabilidad de elegir un punto en cada una de las secciones es la misma, entonces tanto la densidad como la precisión se mejoran en regiones prometedoras del espacio de búsqueda. Refiérase a [9] para detalles sobre la implementación y pormenores del algoritmo.

Como se puede apreciar, pese a que se han hecho algunas aportaciones en el ámbito no paramétrico, aún permanece intocable el área de los algoritmos no paramétricos con dependencias. Una extensión de este tipo en los algoritmos vistos arriba implicaría un enorme incremento del número de cálculos y aumento en la complejidad de los algoritmos. En el siguiente capítulo se hablará sobre una nueva aportación en el ámbito de los algoritmos de estimación de distribución no paramétricos que toman en cuenta las dependencias entre todas las variables.

2.3. Conclusiones del capítulo 2

En el presente capítulo se han definido los términos probabilísticos principales que se usarán a lo largo de este trabajo de investigación, se mostraron algunas características de las funciones de distribución y densidad y su definición tanto para una como para múltiples variables. Asimismo se han revisado los

2.3. Conclusiones del capítulo 2

principales Algoritmos de Estimación de Distribución tanto en su representación discreta como en la continua. Se han presentado los modelos que no realizan suposiciones sobre la distribución sino que la aproximan mediante histogramas marginales, con lo cual éstos no toman en cuenta las dependencias entre las variables. Todos estos modelos y definiciones serán una base informativa para la introducción de dos algoritmos nuevos que toman en cuenta las dependencias entre las variables y no realizan suposiciones previas sobre la distribución de los datos. Uno de estos algoritmos se presenta en el siguiente capítulo.

Capítulo 3

Algoritmo de Estimación de Distribución no paramétrico

En el presente capítulo se presentará un nuevo Algoritmo de Estimación de Distribución el cual toma en cuenta las dependencias entre todas las variables y no hace ninguna suposición previa sobre su distribución conjunta. En su lugar, regenera cada distribución univariada de forma empírica y a cada una de ellas les induce la correlación conjunta de las variables de la muestra mediante un método propuesto por Iman y Conover (1982). Como se verá en las siguientes secciones, el método es libre de una distribución particular y preserva la forma exacta de las distribuciones marginales de entrada, además, como el algoritmo está basado en un método no paramétrico, no utiliza la matriz de covarianzas Σ , sino que emplea la matriz de coeficientes de correlación de Spearman como medida de asociación entre las variables.

Para entrar de lleno en el estudio de este nuevo modelo, se inicia el capítulo con las bases teóricas necesarias para la exposición del algoritmo, se definen y describen las medidas de asociación utilizadas tanto para relaciones lineales como para relaciones no lineales monótonas. Se describe el método de Iman y Conover para la inducción de la distribución y el método para la regeneración de las distribuciones marginales empíricas así como la descripción del algoritmo no paramétrico en su totalidad. Finalmente se muestran las conclusiones y los datos más relevantes del capítulo.

3.1. Medidas de asociación

Como se había mencionado en el capítulo 2, el coeficiente de correlación es una medida normalizada que indica la relación entre dos variables, de tal suerte que su valor absoluto revela la fuerza de la relación y el signo expone si ambas variables crecen a la par o bien, si mientras una crece la otra decrece. Se reserva el término de *correlación* al coeficiente de correlación de Pearson, el cual se

expresa en la ecuación 2.8. Este coeficiente indica las relaciones lineales entre las variables en cuestión, pero ¿qué ocurriría si las relaciones entre las variables fueran no lineales?, en este caso, el coeficiente de correlación de Pearson podría indicar una relación de cero pese a que la existencia de esta relación fuera muy fuerte. Considere como ejemplo que se tiene una variable aleatoria X distribuida uniformemente en $(0,1)$. El coeficiente de correlación de Pearson entre los conjuntos X y X^3 es próximo al 90% pese a que los datos se encuentran determinísticamente relacionados, esto es, que la relación entre ellos es del 100%.

En la literatura se pueden encontrar diversas medidas de asociación que no están limitadas a relaciones lineales y que se basan en el concepto de *rango*. Entre las más populares se pueden nombrar a los coeficientes ρ_S de Spearman y τ de Kendall.

La medida de relación de Spearman para dos variables X e Y está dada por la ecuación 3.1.

$$\rho_S(X, Y) = \frac{\sum_i [R(X_i) - \bar{R}_X] [R(Y_i) - \bar{R}_Y]}{\sqrt{\sum_i [R(X_i) - \bar{R}_X]^2 \sum_i [R(Y_i) - \bar{R}_Y]^2}} \quad (3.1)$$

donde el índice i identifica las instancias x_i e y_i de las variables aleatorias X e Y . $R(X_i)$ es el rango de la i -ésima instancia de la variable X (es decir, el orden de la observación en el conjunto; el rango de la máxima observación es 1, la segunda más grande es 2, y así sucesivamente) y \bar{R}_X es la esperanza del rango de X . El uso de rangos en el cálculo de esta medida de asociación hacen que se conozca también como *coeficiente de correlación por rangos*.

El coeficiente de correlación por rangos de Spearman es una medida de asociación que permite identificar relaciones entre las variables tanto lineales como no lineales siempre que sean monótonas. Una relación monótona puede ser descrita como *monótona creciente* (la cual está asociada con valores positivos de la correlación) o *monótona decreciente* (la cual está asociada con correlaciones negativas). Conforme el valor absoluto del coeficiente de correlación de Spearman se aproxima a uno, la fuerza de la relación monótona aumenta alcanzando su valor máximo cuando ρ_S es $+1$ o bien cuando ρ_S es -1 . Conforme el valor de ρ_S se acerca a cero la relación entre las dos variables es más débil llegando a su mínimo cuando ρ_S es exactamente igual a cero y entonces ninguna relación monótona está presente entre las dos variables. Al igual que en el caso del coeficiente de correlación de Pearson, una correlación positiva indica que el aumento (o decremento) en el valor de una de las variables está asociado con el aumento (o decremento) en el valor de la otra variable, consecuentemente, una correlación negativa indica que el aumento (o decremento) de una de las variables está asociado con el decremento (o aumento) en el valor de la otra variable [10].

El coeficiente de correlación por rangos de Kendall, por otro lado, se encuentra definido por:

$$\tau(X, Y) = 1 - \frac{4}{n(n-1)} \sum_{i=1}^n \sum_{j>i} A_{i,j} \quad (3.2)$$

donde $A_{i,j} = \text{sgn}[(X_i - X_j)(Y_i - Y_j)]$ y sgn es una función que regresa un +1 si el argumento es mayor que cero, -1 si el argumento es menor a cero y 0 en otro caso.

Al igual que la medida de asociación de Spearman, la medida de asociación τ de Kendall, determina relaciones monótonas entre las variables y detecta efectos similares que la ρ_s de Spearman en la población, sin embargo, cuando una muestra es derivada de una distribución normal bivariada, el valor calculado de la ρ_s de Spearman generalmente provee una aproximación razonablemente buena de la correlación de Pearson, mientras que la τ de Kendall no lo hace así. Una ventaja de la τ de Kendall es que a diferencia de la ρ_s de Spearman, la distribución de la muestra de τ se aproxima muy rápidamente a la normal. Debido a esto, la distribución de la muestra de τ puede ser muy bien aproximada por una distribución normal, incluso con muestras muy pequeñas. En contraste, se requiere de una muestra más grande para aproximar la distribución normal a la distribución exacta de la muestra de ρ_s [10].

Tanto la medida de asociación de Spearman como la medida de asociación de Kendall están basadas en el concepto de *concordancia*, la cual se refiere a la propiedad de que valores grandes de una variable aleatoria están asociadas con valores grandes de la otra variable, mientras que la *discordancia* se refiere a que valores grandes de una de las variables está asociado con valores pequeños de la otra variable, con lo cual, la medida τ de Kendall, por ejemplo, se encuentra definida como sigue:

$$\tau(XY) = p[(X_1 - X_2)(Y_1 - Y_2) > 0] - p[(X_1 - X_2)(Y_1 - Y_2) < 0] \quad (3.3)$$

Ya que en los elementos concordantes se cumple que $(X_1 - X_2)(Y_1 - Y_2) > 0$ y en los elementos discordantes se cumple que $(X_1 - X_2)(Y_1 - Y_2) < 0$, entonces:

$$\tau(XY) = p[\text{Concordantes}] - p[\text{Discordantes}] \quad (3.4)$$

A diferencia del coeficiente de correlación de Pearson, las medidas de asociación de Spearman y Kendall son ambas invariantes con respecto a transformaciones estrictamente crecientes y además son iguales a +1 si una de sus variables es una transformación creciente de la otra o bien, iguales a -1 si una de sus variables es la transformación decreciente de la otra. Estas propiedades convierten a ρ_s y a τ como medidas de asociación apropiadas para distribuciones multivariadas no paramétricas [11].

3.2. Inducción de la correlación

En Algoritmos de Estimación de Distribución el principal problema es aproximar la distribución conjunta de la muestra para inducirla a la siguiente generación. Una forma de conseguir este efecto, alterna a las técnicas revisadas en el capítulo dos, es regenerar la distribución de cada variable de forma marginal e inducir la correlación de las variables de la muestra a la nueva población mediante alguna técnica específica. Iman y Conover (1982) proponen un método que permite inducir una cierta matriz de correlación a un conjunto de variables aleatorias de entrada [12]. Este método es simple de usar, se encuentra libre de distribuciones y preserva la forma exacta de las distribuciones marginales de las variables de entrada.

El método está basado en la premisa de que la correlación por rangos es una forma significativa de definir las dependencias entre las variables de entrada. Esto es, que el coeficiente de correlación (Pearson) calculado mediante datos en crudo puede perder el significado y la interpretación de los datos si éstos no se distribuyen de forma normal o con la presencia de datos atípicos. Por otro lado, los coeficientes de correlación por rangos pueden ser más significativos en la mayoría de las situaciones de modelado, incluso cuando los datos provienen de distribuciones normales.

La base teórica del método es que es posible obtener un vector transformado XP' el cual tiene una matriz de correlación deseada C , siempre que los elementos del vector de entrada X estén decorrelacionados y la matriz de correlación C pueda ser escrita como $C = PP'$, con P como una matriz triangular inferior (Scheuer y Stoller, 1962).

Para ejemplificar el método, supóngase que se tienen n vectores de entrada, cada uno con tamaño N y que R es una matriz de tamaño $N \times n$, cuyas columnas representan n permutaciones independientes de un conjunto arbitrario $\{a(i)\}$, $i = 1, \dots, N$ de N números referidos como *puntuaciones*. Cada fila de R (R_i), tiene n componentes independientes, donde cada componente toma uno de los valores $a(i)$ con igual probabilidad. Sea C^* la matriz de correlación deseada y establezca $C = C^*$. Sea P una matriz que cumple $PP' = C$. Es fácil ver que la multiplicación R_iP' resulta en un vector cuyos elementos se relacionan de acuerdo con la matriz de correlación deseada C y que el producto de RP' genera una matriz R^* cuyas filas tienen la misma distribución multivariada que R_iP' . Con lo cual, la matriz de correlación M de R^* se acerca a C .

Finalmente, los valores de cada columna de la matriz R deben ser reacomodados de tal forma que tengan el mismo orden que las correspondientes columnas de R^* . Con ello se garantiza que tanto R como R^* tengan la misma matriz de correlación C^* .

Es verificable el hecho de que esta técnica permite inducir la correlación de manera existosa a un vector o a un grupo de vectores cuando la función de distribución conjunta sea el resultado del producto espacial de las variables in-

volucradas. De no ser así, esta técnica deberá ser reemplazada por otra [13].

Para ilustrar el procedimiento anterior considérese el siguiente ejemplo:

Ejemplo 3.1. *Suponga que la matriz de correlación deseada C^* está dada por:*

$$C^* = \begin{bmatrix} 1.00 & 0.75 & 0.64 & -0.18 & -0.47 \\ 0.75 & 1.00 & 0.46 & -0.13 & -0.20 \\ 0.64 & 0.46 & 1.00 & -0.22 & -0.68 \\ -0.18 & -0.13 & -0.22 & 1.00 & 0.41 \\ -0.47 & -0.20 & -0.68 & 0.41 & 1.00 \end{bmatrix}$$

y que su correspondiente matriz triangular superior P' es obtenida mediante la factorización Cholesky de C^* :

$$P' = \begin{bmatrix} 1.00 & 0.75 & 0.64 & -0.18 & -0.47 \\ 0.00 & 0.65 & -0.03 & 0.00 & 0.24 \\ 0.00 & 0.00 & 0.76 & -0.13 & -0.48 \\ 0.00 & 0.00 & 0.00 & 0.97 & 0.26 \\ 0.00 & 0.00 & 0.00 & 0.00 & 0.63 \end{bmatrix}$$

La matriz R , cuya correlación se desea que coincida con la matriz de correlaciones C^* es obtenida a partir de un vector aleatorio de tamaño $N = 10$ permutado $n = 5$ veces.

$$R = \begin{bmatrix} 1.75 & 0.43 & 0.43 & 4.38 & 2.63 \\ 1.31 & 0.87 & 3.50 & 1.31 & 0.87 \\ 2.63 & 1.31 & 2.63 & 1.75 & 1.75 \\ 3.94 & 1.75 & 1.75 & 3.07 & 1.31 \\ 3.07 & 2.19 & 3.07 & 0.43 & 0.43 \\ 4.38 & 2.63 & 2.19 & 3.50 & 3.07 \\ 3.50 & 3.07 & 3.94 & 0.87 & 4.38 \\ 0.87 & 3.50 & 1.31 & 2.63 & 2.19 \\ 0.43 & 3.94 & 4.38 & 2.19 & 3.50 \\ 2.19 & 4.38 & 0.87 & 3.94 & 3.94 \end{bmatrix}$$

De acuerdo con el método de Iman y Conover, la matriz R^* se obtiene como el producto de RP' .

3.2. Inducción de la correlación

$$R^* = \begin{bmatrix} 1.75 & 1.61 & 1.45 & 3.88 & 3.44 \\ 1.31 & 1.56 & 3.48 & 0.56 & -0.74 \\ 2.63 & 2.85 & 3.65 & 0.86 & -0.14 \\ 3.94 & 4.13 & 3.82 & 2.01 & 0.55 \\ 3.07 & 3.75 & 4.24 & -0.55 & -1.88 \\ 4.38 & 5.04 & 4.41 & 2.30 & 1.41 \\ 3.50 & 4.66 & 5.16 & -0.32 & -0.30 \\ 0.87 & 2.95 & 1.43 & 2.23 & 2.71 \\ 0.43 & 2.90 & 3.47 & 1.47 & 1.78 \\ 2.19 & 4.52 & 1.92 & 3.33 & 4.23 \end{bmatrix}$$

La matriz de correlación por rangos de Spearman M de R^* puede compararse con la matriz de correlaciones deseada C^* .

$$M = \begin{bmatrix} 1.00 & 0.69 & 0.80 & -0.10 & -0.33 \\ 0.69 & 1.00 & 0.56 & 0.06 & 0.06 \\ 0.80 & 0.56 & 1.00 & -0.57 & -0.69 \\ -0.10 & 0.06 & -0.57 & 1.00 & 0.91 \\ -0.33 & 0.06 & -0.69 & 0.91 & 1.00 \end{bmatrix}$$

Puede observarse que pese a M se aproxima mucho a C^* algunos valores son muy diferentes uno del otro. La principal razón de este fenómeno es que la correlación de la matriz R no es exactamente la identidad, por lo cual, existe cierta correlación entre las variables de entrada. En otras palabras, si se cumpliera que la matriz de correlación por rangos de R fuera exactamente la identidad, entonces se cumpliría que $M = C^*$. Obsérvese la matriz de correlación por rangos de R a continuación:

$$T = \begin{bmatrix} 1.00 & -0.15 & -0.06 & -0.02 & -0.02 \\ -0.15 & 1.00 & 0.12 & 0.04 & 0.61 \\ -0.06 & 0.12 & 1.00 & -0.79 & 0.01 \\ -0.02 & 0.04 & -0.79 & 1.00 & 0.34 \\ -0.02 & 0.61 & 0.01 & 0.34 & 1.00 \end{bmatrix}$$

Finalmente, para que la matriz R tenga la misma correlación que R^* , se reordenan los elementos de R de acuerdo con las posiciones de los elementos de R^* . A manera de ilustración se mostrará el ordenamiento de la segunda columna de R :

3.3. Descripción del algoritmo

#	$R_{i,2}$	#	$R_{i,2}^*$		#	$R_{i,2}$
1	0.43	2	1.61		2	0.87
2	0.87	1	1.56		1	0.43
3	1.31	3	2.85		3	1.31
4	1.75	8	4.13		8	3.50
5	2.19	6	3.75	\implies	6	2.63
6	2.63	10	5.04	Ordenando	10	4.38
7	3.07	9	4.66		9	3.94
8	3.50	5	2.95		5	2.19
9	3.94	4	2.90		4	1.75
10	4.38	7	4.52		7	3.07

El símbolo # indica la posición que tendrían los elementos de $R_{i,2}$ y de $R_{i,2}^*$ en caso de estar ordenados de forma creciente. Como se puede apreciar arriba, la segunda columna de R ha quedado ordenada de acuerdo con lo indicado por las posiciones de los elementos de $R_{i,2}^*$.

El uso de esta técnica es primordial en el desarrollo del Algoritmo de Estimación de Distribución no paramétrico y es su principal base teórica. La descripción del mismo y un ejemplo del uso de este modelo se presentan en la siguiente sección.

3.3. Descripción del algoritmo

El Algoritmo de Estimación de Distribución no paramétrico (**NOPREDA**) consta de dos etapas para aproximar la distribución de los mejores individuos e inducirlos a la nueva generación. En la primera de ellas, la distribución marginal de cada variable es capturada y remuestreada para generar las nuevas variables de forma independiente. En la segunda etapa se calcula la correlación de la distribución conjunta y se induce a la nueva población mediante la técnica de Iman y Conover.

El algoritmo 3 muestra la forma general de implementar el NOPREDA.

Hasta el punto 5 el algoritmo del NOPREDA se desarrolla como cualquier otro Algoritmo de Estimación de Distribución. En el punto 6 se efectúa la regeneración de las variables a partir de las distribuciones marginales de las variables de la muestra. Este proceso consiste en generar la distribución marginal empírica $\hat{F}_{X_j}(x_{i,j})$ de cada variable aleatoria j mediante la expresión 3.5.

$$\hat{F}_{X_j}(x_{i,j}) = \frac{f_{\text{acum}}(x_{i,j}) - d}{f_{\text{acum}}(x_{N,j})} \quad (3.5)$$

Donde $x_{i,j}$ representa al valor que toma la j -ésima variable aleatoria del i -ésimo individuo de la población y $f_{\text{acum}}(x_{i,j})$ es la frecuencia acumulada de la

Algoritmo 3 NOPREDA

- 1: $t=0$
- 2: Generar N individuos de forma aleatoria y guardarlos en $P(t)$.
- 3: **while** no se cumpla la condición de paro **do**
- 4: Evaluar cada uno de los individuos de $P(t)$.
- 5: Seleccionar $M \leq N$ individuos de $P(t)$ y colocarlos en la muestra S .
- 6: **for** $j = 1$ a n **do**
- 7: Regenerar la j -ésima variable a partir de la distribución marginal de la variable j de S y guardarla en $P(t + 1)$.
- 8: **end for**
- 9: Inducir la correlación conjunta de la muestra S en $P(t + 1)$.
- 10: $t=t+1$
- 11: **end while**

j -ésima variable aleatoria hasta el i -ésimo elemento de la variable aleatoria X_j . Asimismo, d es una constante ($d \in (0, 1)$, comúnmente $d=0.5$) que permite la interpolación entre los datos discretizados, observe la figura 3.1. Sin este ajuste, existirían instancias con probabilidad de elección igual a cero, eliminando así su capacidad de regeneración en futuras generaciones.

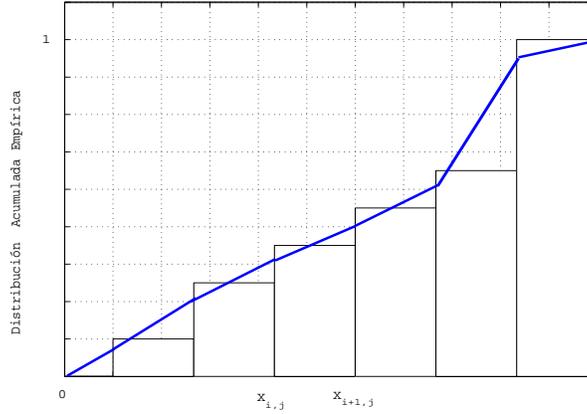


Figura 3.1: Interpolación de los datos para la estimación de la distribución.

En la figura 3.1 se puede apreciar el histograma de frecuencias acumuladas compuesto por cada banda de frecuencias $b_i = x_{i+1,j} - x_{i,j}$ y el diagrama de la distribución acumulada empírica, la cual suaviza los valores del histograma y permite la generación de valores interpolados entre $x_{i,j}$ y $x_{i+1,j}$.

El procedimiento para la obtención de cada distribución acumulada es como sigue: primero se ordenan los elementos de la j -ésima variable aleatoria de forma creciente incluyendo los límites de las variables y suprimiendo los valores repetidos, llámase a este vector X_j^c y a partir de éste obténgase las correspondientes

3.3. Descripción del algoritmo

frecuencias acumuladas para cada instancia de este vector aleatorio mediante la ecuación 3.5. Enseguida constrúyase la gráfica acumulada para estos valores de X_j^e como se muestra en la figura 3.2.

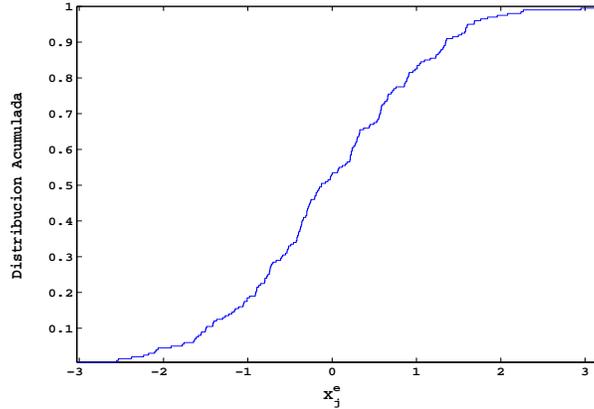


Figura 3.2: Frecuencia acumulada empírica de la j -ésima variable

Luego de la construcción de la gráfica, se regenera cada variable mediante el muestreo de N números aleatorios uniformes $(0,1)$ para cada instancia i de la j -ésima variable y se evalúan en la inversa de la función de distribución acumulada empírica, como se ilustra en la figura 3.3.

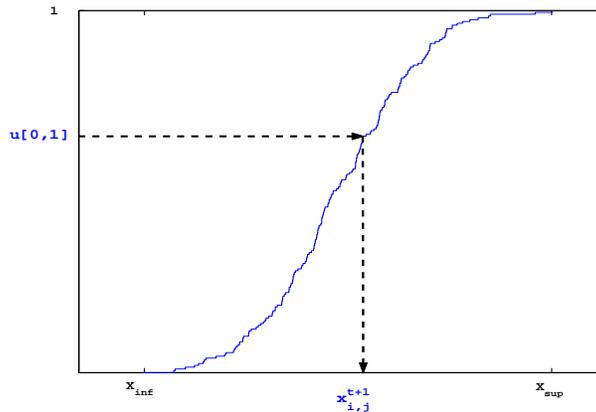


Figura 3.3: Muestreo aleatorio para la regeneración de la i -ésima instancia de la variable j .

Este proceso involucra la interpolación de los valores del vector X_j^e de forma que siempre se estén generando nuevas instancias del vector aleatorio. El tipo de interpolación a emplear para el muestreo de las variables no queda restringido

3.3. Descripción del algoritmo

sólo a lineal o a polinomial, si no que pueden emplearse un sinnúmero de técnicas dependiendo del tamaño de la población utilizada o la finura que se desea en la generación de las instancias. A manera de ilustración, se indica en la ecuación 3.6 la forma de realizar la interpolación lineal para generar la nueva instancia $x_{i,j}^{t+1}$.

$$x_{i,j}^{t+1} = x_{i,j}^{e,t} + \frac{(u - F_{X_j}(x_{i,j}^{e,t})) (x_{i+1,j}^{e,t} - x_{i,j}^{e,t})}{F_{X_j}(x_{i+1,j}^{e,t}) - F_{X_j}(x_{i,j}^{e,t})} \quad (3.6)$$

donde $x_{i,j}^{e,t}$ representa el valor de la i -ésima instancia del vector X_j^e en la generación t y u representa un valor uniforme (0,1) empleado para la regeneración de la i -ésima instancia del vector j . El cálculo para la interpolación lineal descrito por la ecuación 3.6 es válido siempre que u se encuentre entre $x_{i,j}^{e,t}$ y $x_{i+1,j}^{e,t}$.

Cuando todas las variables se han regenerado, se tendrá un conjunto de variables que tienen una distribución marginal semejante a las variables de la muestra. Como es deseable que la relación entre todas las variables de la muestra se encuentre implícita en la nueva población, se emplea entonces el procedimiento de Iman y Conover descrito en la sección anterior. Este proceso se indica en el paso 8 del algoritmo 3. Para coincidir con el método de inducción de correlaciones se genera una matriz de permutaciones R a partir de un vector aleatorio uniforme (0,1). La matriz de correlación deseada C^* es la misma matriz de correlación por rangos de Spearman de los elementos de la muestra y ésta es la que se factoriza mediante Cholesky. Finalmente el orden de los elementos de la matriz $R^* = RP'$ indica el orden al que se deberán reordenar los elementos de la nueva población generada de forma marginal empírica.

En el siguiente ejemplo se muestran los efectos de emplear el método de regeneración con normales multivariadas y con NOPREDA para una población inicial dada.

Ejemplo 3.2. *Considere la siguiente población inicial:*

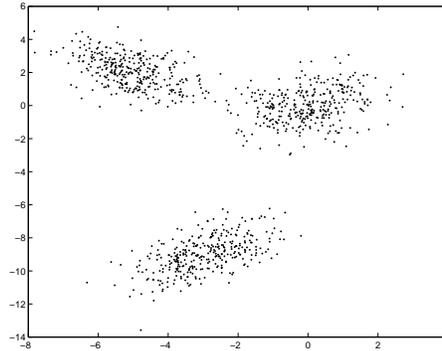


Figura 3.4: Población inicial con coeficiente $\rho_S = -0.3143$

3.3. Descripción del algoritmo

La cual está formada por 1000 datos provenientes de tres gaussianas con vectores de medias $\boldsymbol{\mu}_1=(-5,2)$, $\boldsymbol{\mu}_2=(-3,-9)$, $\boldsymbol{\mu}_3=(0,0)$ y matrices de covarianzas $\Sigma_1=[1 \ -0.5; -0.5 \ 1]$, $\Sigma_2=[1 \ 0.6; 0.6 \ 1]$, $\Sigma_3=[1 \ 0.2; 0.2 \ 1]$.

Si la población inicial es regenerada mediante normales multivariadas se obtiene el siguiente resultado:

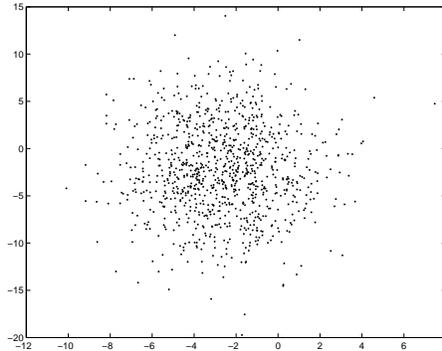


Figura 3.5: Población regenerada mediante Normales Multivariadas $\rho_S = -0.0055$

Como se puede observar en la imagen anterior la forma de la población regenerada no se aproxima a la población inicial y además los coeficientes de correlación por rangos son totalmente diferentes, lo cual muestra que este proceso no es precisamente eficiente para preservar las dependencias entre las variables originales.

Si ahora se compara con la población regenerada mediante NOPREDA, mostrada en la figura 3.6, se puede ver que en este caso la forma de los datos regenerados se aproxima a los originales así como los coeficientes de correlación por rangos.

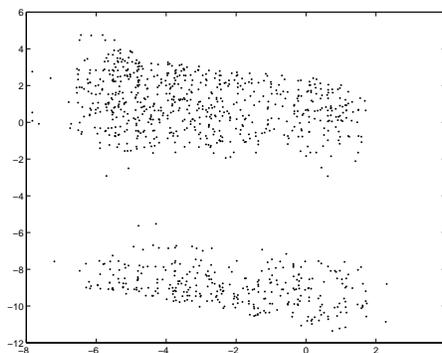


Figura 3.6: Población regenerada mediante NOPREDA $\rho_S = -0.3096$

En el ejemplo anterior se pudo observar que el procedimiento del NOPREDA puede preservar las correlaciones iniciales de un conjunto de datos de forma

más precisa que una técnica de regeneración gaussiana multivariada, lo cual la destaca como una herramienta útil en el proceso de búsqueda de los Algoritmos de Estimación de Distribución.

3.4. Conclusiones del capítulo 3

En el presente capítulo se presentó un algoritmo de Estimación de Distribución No Paramétrico conocido como NOPREDA, el cual emplea el método de Iman y Conover para inducir la correlación conjunta de la muestra hacia los nuevos individuos de la población. Primeramente, se habló de los métodos para medir la asociación entre las variables y se establecieron las medidas que se pueden emplear en el método del NOPREDA debido a su flexibilidad para indicar no sólo relaciones lineales entre las variables sino también monótonas. Posteriormente se explicó el método de Iman y Conover para la inducción de una matriz de correlación deseada hacia un conjunto de datos de entrada para luego entrar de lleno en la explicación del método del NOPREDA. Como se vio, este método regenera cada variable a partir de la distribución marginal empírica de la muestra y luego al nuevo conjunto así obtenido, se le induce la correlación conjunta deseada de los individuos seleccionados.

Capítulo 4

Algoritmo de Estimación de Distribución con Cópulas

En este capítulo se presenta un nuevo algoritmo de estimación de distribución el cual emplea una cópula gaussiana (COPULEDA, por sus siglas) para inducir la correlación de las variables de la muestra hacia la nueva población. La noción de *cópula* fue introducida por primera vez por A. Sklar en 1959, aunque la idea ya había aparecido con anterioridad en algunos textos y con mayor notoriedad en el de Hoeffding (1940, 1941). Sklar acuñó el término de cópula para designar aquella función que relaciona la función de probabilidad multidimensional con sus funciones marginales de bajo orden. Esta función está restringida a $[0, 1]^n$ y sus marginales están en $[0, 1]$. Como se verá en las siguientes secciones, Sklar demostró que si H es una función de distribución con marginales $F(x)$ y $G(y)$, entonces existe una cópula C tal que $H(x, y) = C(F(x), G(y))$.

Al principio, las cópulas fueron empleadas para el desarrollo de la teoría de la métrica estadística, posteriormente, las cópulas se emplearon para definir medidas de asociación no paramétrica entre las variables y desde entonces se utilizan para resolver una gran diversidad de problemas en la rama estadística matemática, actuarial, económica y financiera, entre otras [14].

En primera instancia, se definirá el término de *cópula* y se presentarán sus principales propiedades. Se hablará sobre las diferentes familias de cópulas y sus características relevantes. Además, se establecerán las equivalencias de las medidas de asociación de acuerdo con el tipo de cópula empleada y se ejemplificará la forma de realizar la simulación de datos a partir de una cópula. Luego de estas bases teóricas, se entrará de lleno en la explicación de algoritmo del COPULEDA para finalmente concluir con los puntos más sobresalientes del capítulo.

4.1. Cópulas: definiciones y propiedades.

Las cópulas son funciones de distribución conjunta las cuales nos permiten separar los efectos de las dependencias entre las variables de los efectos de sus correspondientes distribuciones marginales. Debido a que las cópulas son funciones de distribución, comparten las mismas propiedades vistas en el capítulo 2. Además de éstas, defínase los siguientes conceptos importantes en la teoría de cópulas:

Definición 4.1 (Función de supervivencia). Sea (Y_1, \dots, Y_n) un conjunto de variables aleatorias y $F_{Y_1, \dots, Y_n}(y_1, \dots, y_n)$ su correspondiente función de distribución conjunta, defínase entonces

$$\bar{F}_{Y_1, \dots, Y_n}(y_1, \dots, y_n) = p\{Y_i > y_i; i = 1, \dots, n\}$$

como la función de supervivencia de $F_{Y_1, \dots, Y_n}(y_1, \dots, y_n)$

Observe que, de acuerdo con la definición 4.1, la función de supervivencia de una función de distribución es sencillamente el complemento de la función de distribución, esto es $\bar{F}(y_1) = 1 - F(y_1)$, $\bar{F}(y_1, y_2) = 1 - F_1(y_1) - F_2(y_2) + F_1(y_1)F_2(y_2)$, etcétera.

Definición 4.2 (Límites de Frèchet-Hoeffding). Sea (Y_1, \dots, Y_n) un conjunto de variables aleatorias y $F_{Y_1, \dots, Y_n}(y_1, \dots, y_n)$ su correspondiente función de distribución conjunta cuyas marginales univariadas están representadas por F_1, \dots, F_n , defínase entonces

$$F_L(y_1, \dots, y_n) = \max \left[\sum_{j=1}^n F_j - n + 1, 0 \right]$$

$$F_U(y_1, \dots, y_n) = \min[F_1, \dots, F_n]$$

como los límites inferior y superior de Frèchet-Hoeffding de $F_{Y_1, \dots, Y_n}(y_1, \dots, y_n)$.

Como se ha dicho anteriormente, las funciones de distribución univariadas se encuentran acotadas por arriba en 1 y por abajo en 0, ahora bien, de acuerdo con la definición 4.2, se puede observar que una función de distribución conjunta se encuentra acotada según los límites inferior y superior de Frèchet-Hoeffding, de tal forma que el mínimo límite inferior es cero y el máximo límite superior está dado por el valor mínimo de las funciones de distribución marginales.

Ahora bien, considérese la siguiente definición de cópula para el caso n -dimensional:

Definición 4.3 (Cópula). Una cópula n -dimensional es una función $C : [0, 1]^n \rightarrow [0, 1]$ la cual satisface:

1. $C(1, \dots, 1, a_m, 1, \dots, 1) = a_m$, para todo $m \leq n$ y toda $a_m \in [0, 1]$.
2. $C(a_1, \dots, a_n) = 0$ si $a_m = 0$, para todo $m \leq n$.
3. C es n -creciente.

La definición 4.3 es conocida como el **teorema de Sklar** y una implicación práctica de dicho teorema es que de existir dos variables aleatorias X e Y con función de distribución conjunta H y funciones de distribución marginales $F(x) = u_1$ y $G(y) = u_2$, respectivamente, existe una cópula C que cumple:

$$H(x, y) = C(F(x), G(y)) \quad (4.1)$$

para todo $x, y \in \mathbb{R}$. Además, trabajando con 4.1 y suponiendo que F y G son continuas, se puede llegar a la siguiente serie de igualdades:

$$\begin{aligned} H(x, y) &= p\{X \leq x, Y \leq y\} \\ &= p\{X \leq F^{-1}(u_1), Y \leq G^{-1}(u_2)\} \\ &= H(F^{-1}(u_1), G^{-1}(u_2)) \end{aligned}$$

lo cual implica que:

$$C(F(x), G(y)) = H(F^{-1}(u_1), G^{-1}(u_2)) \quad (4.2)$$

y de acuerdo con el teorema de Sklar, ésta es la única solución de 4.1. Si no se cumple que F y G son continuas, o bien si sólo una de ellas lo es, la cópula C estaría únicamente determinada en $\text{Rango}(F) \times \text{Rango}(G)$, esto significa que la función de distribución conjunta siempre podrá ser expresada como 4.1, aunque en este caso la cópula no será única [15].

Ya que las cópulas son funciones de distribución multivariadas, los límites de Frèchet-Hoeffding también son válidos para ellas. Se denotará por C_L al límite inferior de Frèchet-Hoeffding de la cópula C y con C_U su límite superior de Frèchet-Hoeffding. Finalmente, el símbolo C^\perp denota que las marginales de la cópula C son independientes. Si una familia de cópulas incluye C_L, C_U y C^\perp se dice que es *completa*.

Algunas propiedades adicionales de las cópulas se enuncian enseguida para el caso $n = 2$:

1. Las variables aleatorias Y_1 y Y_2 son independientes sí y sólo si se cumple $C(F_1(y_1), F_2(y_2)) = F_1(y_1)F_2(y_2)$.
2. Y_1 es una función creciente de Y_2 sí y sólo si $C(\cdot) = C_U(\cdot)$.
3. Y_1 es una función decreciente de Y_2 sí y sólo si $C(\cdot) = C_L(\cdot)$.
4. Las cópulas poseen la propiedad de invariancia, lo cual significa que la dependencia capturada por una cópula es invariante con respecto a transformaciones crecientes y continuas de las distribuciones marginales.

Es fácil demostrar la propiedad 1 si se considera independencia entre las variables Y_1, Y_2 y se hace uso de la expresión 4.1. Obsérvese el proceso siguiente:

$$\begin{aligned} C(F_1(y_1), F_2(y_2)) &= F(y_1, y_2) && \text{teorema de Sklar} \\ &= F_1(y_1)F_2(y_2) && \text{por independencia} \end{aligned}$$

con lo cual se demuestra la propiedad 1. Una implicación importante de la propiedad 4 es que la estructura de dependencia de la función de distribución conjunta está determinada por la la matriz de correlación inherente a esa distribución [16]. Algunos comentarios importantes sobre las propiedades restantes pueden encontrarse en [17].

En el siguiente ejemplo se muestra la forma de calcular la cópula asociada a una función de distribución conjunta dada empleando el teorema de Sklar.

Ejemplo 4.1. Sean X e Y dos variables aleatorias continuas, $H(x, y)$ su correspondiente función de distribución conjunta y $F(x)$ y $G(y)$ sus funciones de distribución marginales. Establézcase:

$$H(x, y) = \begin{cases} \frac{(x+1)(e^y-1)}{x+2e^y-1}, & (x, y) \in [-1, 1] \times [0, \infty], \\ 1 - e^{-y}, & (x, y) \in (1, \infty] \times [0, \infty], \\ 0, & \text{otro caso,} \end{cases} \quad (4.3)$$

y sus marginales $F(x)$ y $G(y)$ como:

$$F(x) = \begin{cases} 0, & x < -1, \\ (x+1)/2, & x \in [-1, 1], \\ 1, & x > 1, \end{cases} \quad y \quad G(y) = \begin{cases} 0, & y < 0, \\ 1 - e^{-y}, & y \geq 0. \end{cases}$$

Para obtener la cópula asociada a la función de distribución $H(x, y)$ se emplea la expresión 4.2 y las pseudoinversas de las funciones de distribución marginales. Estas pseudoinversas son equivalentes a las inversas de $F(x)$ y $G(y)$ sólo cuando x e y están entre 0 y $F(0)$ o $G(0)$, respectivamente, con lo cual, las pseudoinversas de F y G están dadas por $F^{[-1]}(u) = 2u - 1$ y $G^{[-1]}(v) = -\ln(1 - v)$, para $u, v \in \mathbf{I} = [0, 1]$. Ya que el $\text{Rango}(F) = \text{Rango}(G)$ y estos son iguales a \mathbf{I} , la ecuación 4.2 nos conduce a la cópula:

$$C(u, v) = \frac{uv}{u + v - uv}.$$

En la figura 4.1 se muestra la gráfica de la distribución conjunta $H(x, y)$ con $(x, y) \in [-1, 1] \times [0, 2]$.

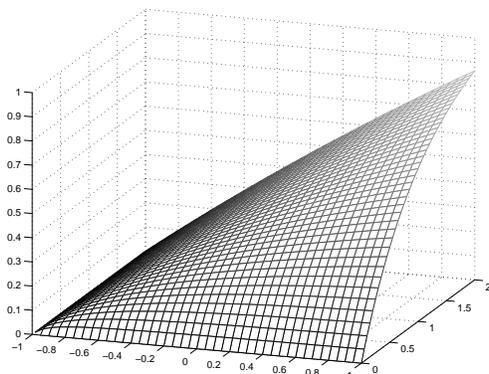


Figura 4.1: Función de distribución conjunta $H(x, y)$

4.1.1. Familias paramétricas de cópulas

Generalícese la ecuación 4.1 como sigue:

$$F(x_1, \dots, x_n) = C(F_1(x_1), \dots, F_n(x_n); \alpha) \quad (4.4)$$

donde α representa el *parámetro de dependencia* de la cópula, y mide la dependencia entre las marginales. Este parámetro o parámetros de dependencia están fuertemente ligados con el tipo de cópula empleada. Existen diversas familias de cópulas que pertenecen asimismo a diversas categorías generales. Las categorías más frecuentemente utilizadas son las *elípticas* y las *arquimedianas*, aunque también existen las cópulas del valor extremo, las cópulas cuadráticas, las Archimax [18], etcétera. A la categoría elíptica [19] pertenecen las familias de cópulas Normal y Student, mientras que entre las arquimedianas se pueden nombrar las cópulas de la familia Gumbel, Clayton, Frank, etcétera.

Existe diversas formas de generar una función de cópula, algunas técnicas conocidas son el método de inversión, el método de Rüschendorf, el método de transformadas de Laplace y mezclas de potencias [11] [18], etcétera. El método de inversión es aplicable a cópulas de distribuciones conocidas como en el caso de la cópula Normal o la Student. Considere el siguiente ejemplo:

Ejemplo 4.2. *Suponga que las variables aleatorias X_1 y X_2 son continuas y se distribuyen de forma normal bivariada estándar con una matriz de correlación lineal $\Sigma = \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}$. Sea Φ_ρ su correspondiente función de distribución conjunta con parámetro de dependencia $\alpha = \rho$ y sea ϕ la función de distribución acumulada univariada estándar. Observe que las funciones de distribución marginales están relacionadas por una función de cópula C única, conocida como *cópula Normal*, que de acuerdo con el teorema de Sklar cumple:*

$$C(u_1, u_2; \rho) = \Phi_\rho(\phi^{-1}(u_1), \phi^{-1}(u_2)) \quad (4.5)$$

4.1. Cópulas: definiciones y propiedades.

A partir de la expresión 4.5 se puede calcular la función de densidad de la cópula $C(u_1, u_2; \rho)$ denotada por $c(u_1, u_2; \rho)$ mediante:

$$\begin{aligned} c(u_1, u_2; \rho) &= \frac{\partial^2 C(u_1, u_2; \rho)}{\partial u_1 \partial u_2} \\ &= \frac{\partial^2 \Phi_\rho [\phi^{-1}(u_1), \phi^{-1}(u_2)]}{\partial u_1 \partial u_2} \cdot \frac{\partial \phi^{-1}(u_1)}{\partial u_1} \cdot \frac{\partial \phi^{-1}(u_2)}{\partial u_2} \end{aligned}$$

Realizando los cálculos, tenemos:

$$\begin{aligned} c(u_1, u_2; \rho) &= \frac{1}{\sqrt{1-\rho^2}} \exp \left\{ -\frac{[\phi^{-1}(u_1)]^2 + [\phi^{-1}(u_2)]^2 - 2\rho\phi^{-1}(u_1)\phi^{-1}(u_2)}{2(1-\rho^2)} \right\} \\ &\quad \times \exp \left\{ \frac{[\phi^{-1}(u_1)]^2 + [\phi^{-1}(u_2)]^2}{2} \right\} \end{aligned} \quad (4.6)$$

Note que un caso especial de la cópula Normal se obtiene si se sustituye $\rho = 0$ en la expresión 4.5, con lo cual $C^\perp(u_1, u_2; \rho) = u_1 u_2$, que es conocida como *cópula producto* o *cópula independiente* y denota independencia entre las variables. Es importante hacer notar ciertas propiedades especiales de las cópulas Normales. Una de ellas es que la cópula normal bivariada denotada por 4.5 tiene extensión al caso multivariado y se extiende también para el caso de dependencias negativas. Además, toma el valor C_U cuando el parámetro de dependencia $\rho = 1$, C^\perp cuando $\rho = 0$ y C_L cuando $\rho = -1$. Esto significa, como se había indicado antes, que la familia Normal es *completa*. En la figura 4.2 se muestran las curvas de nivel de la función de densidad de la cópula Normal expresada en 4.6.

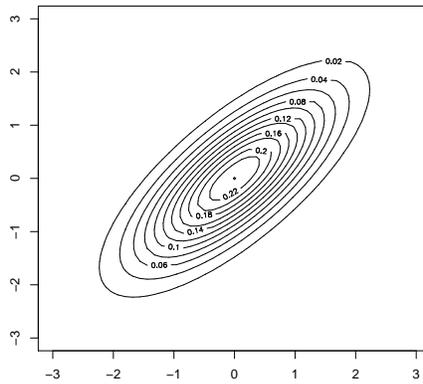


Figura 4.2: Curvas de nivel de la función de densidad de una cópula Normal con parámetro $\alpha = 0.75$ y marginales normales.

La técnica de las transformadas de Laplace y mezclas de potencias de funciones de distribución univariadas o de funciones de supervivencia es útil para generar funciones de distribución multivariadas y con ello, cópulas. Observe el siguiente ejemplo:

Ejemplo 4.3. Sea M la función de distribución univariada de una variable aleatoria no negativa (esto es que $M(0) = 0$) y φ su correspondiente transformada de Laplace, esto es:

$$\varphi(s) = \int_0^{\infty} e^{-sw} dM(w), \quad s \geq 0.$$

Para una función de distribución univariada arbitraria F , existe una única función de distribución acumulada G tal que:

$$F(x) = \int_0^{\infty} G^{\alpha}(x) dM(\alpha) = \varphi(-\log G(x)). \quad (4.7)$$

Reescribiendo 4.7 se conduce a $G = \exp\{-\varphi^{-1}(F)\}$. Ahora bien, considere la clase bivariada formada por F_1, F_2 y sea $G_j = \exp\{-\varphi^{-1}(F_j)\}$ con $j = 1, 2$. Entonces, la siguiente es la función de distribución acumulada bivariada con F_1, F_2 como sus marginales:

$$\int_0^{\infty} G_1^{\alpha} G_2^{\alpha} dM(\alpha) = \varphi(-\log G_1 - \log G_2) = \varphi(\varphi^{-1}(F_1) + \varphi^{-1}(F_2)). \quad (4.8)$$

La cópula que se obtiene a partir de 4.8, se logra sustituyendo los valores $U(0, 1)$ en las correspondientes funciones de distribución univariadas F_1, F_2 , como sigue:

$$C_{\varphi}(u_1, u_2) = \varphi(\varphi^{-1}(u_1) + \varphi^{-1}(u_2)). \quad (4.9)$$

La expresión 4.9 es la forma sencilla de un cópula arquimediana.

En el ejemplo anterior se introduce una clasificación de cópulas conocidas como *cópulas arquimedianas* las cuales se expresan de forma general mediante 4.9 y las cuales deben su nombre al axioma arquimediano de los números reales positivos [20].

Como se indicó en ejemplo 4.3, M corresponde a la función de distribución acumulada de una variable aleatoria no negativa y φ por definición, tiene la propiedad de ser continua y estrictamente decreciente. Asimismo, como $\varphi(0) = 1$, entonces $\varphi^{-1}(0) = \infty$ y $\varphi^{-1}(1) = 0$.

Todos estos elementos nos pueden llevar a una forma más común de definir a la cópula arquimediana, indicada por la ecuación 4.10:

$$C_{\psi}(u_1, u_2) \equiv \psi^{[-1]}(\psi(u_1) + \psi(u_2)) \quad (4.10)$$

en la cual $\psi : [0, 1] \rightarrow \mathbb{R}^+$ es una función decreciente, continua y convexa, tal que $\psi(1) = 0$ y $\psi^{[-1]}$ es la *pseudoinversa* de ψ definida como:

$$\psi^{[-1]}(t) = \begin{cases} \psi^{-1}(t), & 0 \leq t \leq \psi(0), \\ 0, & \psi(0) \leq t \leq \infty. \end{cases} \quad (4.11)$$

Si $\psi(0) = \infty$, como en el caso del ejemplo anterior, entonces $\psi^{[-1]} = \psi^{-1}$ y se dice que ψ es *estricto*.

Pese a que tanto 4.9 como 4.10 son equivalentes al emplear $\psi(\cdot) = \varphi^{-1}(\cdot)$, $\psi(\cdot)^{[-1]} = \psi(\cdot)^{-1} = \varphi(\cdot)$, la expresión 4.10 es más frecuente hallarla en la literatura y su empleo en este trabajo permite unificar los valores de ψ .

En las cópulas arquimedianas, ψ se conoce como el *generador* de la cópula y define una gran variedad de *familias* arquimedianas. Algunos ejemplos de familias de cópulas arquimedianas y sus generadores se muestran a continuación:

1. **Familia Clayton.** Esta familia es construida a partir del generador $\psi(t) = t^{-\alpha} - 1$ con $\alpha > 0$. Esta familia se caracteriza porque conforme el parámetro α tiende a infinito, la dependencia entre las variables se maximiza y conforme tiende a uno, el par (U, V) se independiza. Además, puede realizarse la extensión para que esta familia tenga dependencias negativas empleando el generador $\psi(t) = 1 - t^{1-\alpha}$ con $0 < \alpha < 1$.
2. **Familia Frank.** En este caso, el generador es $\psi(t) = \ln \frac{e^{\alpha t} - 1}{e^{\alpha} - 1}$, con $0 \geq \alpha \geq 1$. En esta familia, la dependencia se hace máxima cuando α tiende a cero y la independencia se alcanza cuando $\alpha = 1$. Al igual que en la familia Clayton, es posible extender esta familia al caso de dependencias negativas cuando $\alpha > 1$.
3. **Familia Gumbel.** Para obtener esta familia el generador es como sigue: $\psi(t) = (-\log(t))^\alpha$, con $\alpha \geq 1$. Conforme el valor de α tiende a infinito, la dependencia se hace máxima, mientras que cuando tiende a cero el par (U, V) se hace independiente. Para esta familia, no existe la extensión al caso de dependencias negativas.
4. **Cópula Independiente.** Su generador es $\psi(t) = -\ln(t)$ y se encuentra definida como antes: $C^\perp(u, v) = uv$.

En las figuras 4.3 a 4.5 se muestran las curvas de nivel de las funciones de densidad de algunas de las familias arquimedianas más comunes.

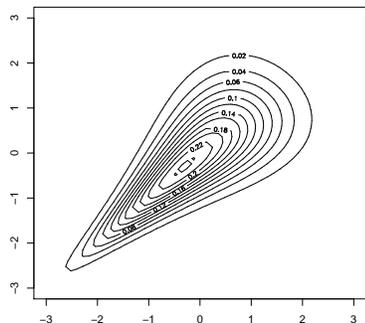


Figura 4.3: Curvas de nivel de la función de densidad de la familia Clayton con parámetro $\alpha = 2$ y marginales normales.

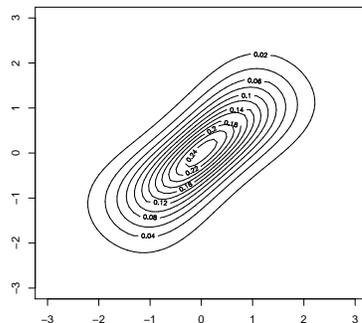


Figura 4.4: Curvas de nivel de la función de densidad de la familia Frank con parámetro $\alpha = 5.736$ y marginales normales.

Es importante hacer notar que las cópulas arquimedianas con dimensión mayor o igual a tres sólo permite asociaciones positivas. Las asociaciones negativas son permitidas en en el caso bivariado, pero en sólo algunas de ellas, como se describió anteriormente.

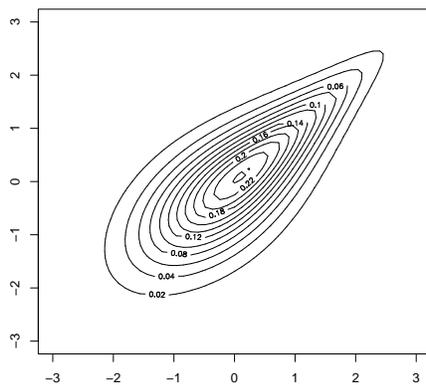


Figura 4.5: Curvas de nivel de la función de densidad de la familia Gumbel con parámetro $\alpha = 2$ y marginales normales.

4.2. Medidas de asociación

En el capítulo anterior se habló sobre algunas medidas de asociación empleadas para el caso de dependencias lineales y no lineales. Establezcamos las propiedades que debe cumplir una medida de asociación:

1. La medida debe estar normalizada para que tome valores entre -1 y +1.
2. Debe ser invariante a escalamientos.
3. Debe tomar el valor de cero si los datos medidos son independientes.
4. Debe crecer conforme la dependencia aumenta y disminuir cuando la dependencia entre los datos disminuye.

Observe que aunque las propiedades 1, 2 y 4 se cumplen para el coeficiente de correlación de Pearson ρ , ya se ha hablado sobre los inconvenientes de esta medida de asociación respecto a la propiedad 3 cuando se involucran relaciones no lineales entre las variables. Es por ello que este trabajo se enfoca en dos medidas de asociación no lineales, las cuales cumplen con las cuatro propiedades mencionadas arriba y además cuentan con algunas características que se enunciarán enseguida. A continuación, se hablará sobre las medidas ρ_s y τ en términos de cópulas.

4.2.1. Coeficiente de correlación por rangos de Kendall

Como se había mencionado antes, la medida de asociación de Kendall se puede definir como la diferencia entre las probabilidades de dos variables aleatorias concordantes y las probabilidades de dos variables aleatorias discordantes. Debido a la propiedad de invarianza ante transformaciones estrictamente crecientes, es posible expresar esta medida de asociación en términos de cópulas de la forma siguiente:

$$\tau(X, Y) = 4 \int_0^1 \int_0^1 C(u, v) dC(u, v) - 1 \quad (4.12)$$

donde $C(u, v)$ es la cópula asociada a (X, Y) . Una demostración de la validez de esta expresión puede encontrarse en [20]. Es fácil ver que esta medida de asociación cumple con las propiedades descritas anteriormente y que:

$$\tau = -1 \quad \text{sí y sólo si } C = C_L \quad (4.13a)$$

$$\tau = 0 \quad \text{sí y sólo si } C = C^\perp \quad (4.13b)$$

$$\tau = 1 \quad \text{sí y sólo si } C = C_U \quad (4.13c)$$

Observe que la integral que aparece en la expresión 4.12, se puede interpretar como el valor esperado de la función $C(u, v)$, esto es:

$$\tau(X, Y) = 4E(C(u, v)) - 1. \quad (4.14)$$

De aquí es fácil ver la relación que existe entre la cópula asociada al par X, Y y la $\tau(X, Y)$. Note que de un conjunto de datos es posible obtener el coeficiente de correlación por rangos de Kendall y de esta forma conocer la relación entre ellos; por otro lado, de conocer la cópula asociada a este par, también es posible conocer la relación entre los datos (medida de asociación), lo cual indica que en efecto, la función de cópula asociada al par X, Y posee de forma implícita las relaciones entre estas variables.

4.2.2. Coeficiente de correlación por rangos de Spearman

Este coeficiente de correlación puede definirse también en términos de concordancia y discordancia como se hace para la medida de asociación de Kendall. Para obtener la definición poblacional de esta medida, sean $(X_1, Y_1), (X_2, Y_2)$ y (X_3, Y_3) tres vectores aleatorios independientes y C la correspondiente cópula asociada. Entonces la versión poblacional de la ρ_s es proporcional a la probabilidad de concordancia menos la probabilidad de discordancia de los dos vectores (X_1, Y_1) y (X_2, Y_3) , lo cual significa que un par de vectores poseen las mismas marginales, pero uno de ellos tiene una distribución conjunta mientras que las componentes del otro son independientes:

$$\rho_s = 3(p[(X_1 - X_2)(Y_1 - Y_3) > 0] - p[(X_1 - X_2)(Y_1 - Y_3) < 0]) \quad (4.15)$$

El 3 que aparece en la expresión 4.15 es una constante de normalización que permite $-1 \leq \rho_s \leq 1$, vea [20].

Al igual que el coeficiente de correlación de Kendall, la ρ_s de Spearman cumple con 4.13a a 4.13c. Redefínase, entonces ρ_s en términos de cópulas como:

$$\rho_s(X, Y) = 12 \int_0^1 \int_0^1 \{C(u, v) - uv\} dudv \quad (4.16)$$

donde $C(u, v)$ es la cópula asociada al par X, Y . Obsérvese la ecuación 4.16. Esta expresión representa la diferencia entre el valor de una probabilidad conjunta $C(u, v)$ y el resultado de una conjunta independiente $C(u, v) = uv$. Este hecho permite identificar a ρ_s como el promedio de la distancia entre las distribuciones X e Y y la independencia Π .

Además de estas expresiones, el coeficiente de correlación de Spearman puede ser visto como el coeficiente de correlación lineal entre dos variables aleatorias uniformes $U = F_X(x)$ y $V = F_Y(y)$, esto es:

$$\rho_s(X, Y) = \rho(F_X(x), F_Y(y)), \quad (4.17)$$

siempre que X e Y sean continuas.

4.2.3. Relaciones entre ρ_s y τ de Kendall

Pese a que ρ_s y τ son ambas medidas de asociación que pueden ser definidas en términos de concordancia y discordancia, los valores de éstas son diferentes, de hecho, las relaciones entre ambas varía de familia a familia. Daniels [24] dió una desigualdad universal que se cumple para estas medidas:

$$-1 \leq 3\tau - 2\rho_s \leq 1 \quad (4.18)$$

cuya demostración puede encontrarse en [25]. Además de esta desigualdad, Durbin y Stuart [23] mostraron que las relaciones entre el coeficiente de correlación de Spearman y la τ de Kendall estaban dadas por las siguientes expresiones:

$$\frac{3\tau - 1}{2} \leq \rho_s \leq \frac{1 + 2\tau - \tau^2}{2} \quad \text{para } \tau \geq 0 \quad (4.19)$$

$$\frac{\tau^2 + 2\tau - 1}{2} \leq \rho_s \leq \frac{1 + 3\tau}{2} \quad \text{para } \tau \leq 0. \quad (4.20)$$

Un bosquejo del comportamiento de estas medidas de asociación en variables aleatorias continuas se muestra en la figura 4.6. Esta imagen indica que para cualquier par de variables aleatorias continuas X, Y , la versión poblacional de la τ de Kendall y de la ρ de Spearman deben encontrarse en cualquier parte de la zona sombreada conocida como la región $\tau - \rho$.

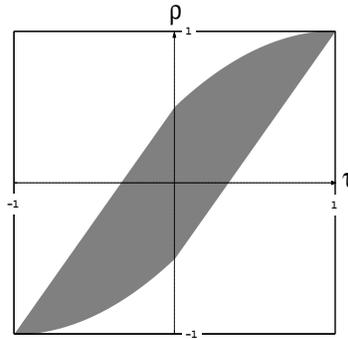


Figura 4.6: Comportamiento de ρ_s y τ para variables aleatorias continuas.

Además de estas relaciones entre los coeficientes de correlación, es posible especificar otras relaciones entre ρ_s, τ y el parámetro de dependencia α para algunas de las cópulas más populares. En el cuadro 4.1 se resumen las características de los parámetros de dependencia y su relación en términos de las medidas de asociación ρ_s y τ .

Para ver otras equivalencias entre los parámetros de dependencia de diversas familias de cópulas bivariadas y las medidas de asociación ρ_s y τ se puede revisar [11],[20].

4.2. Medidas de asociación

Cópula	Dominio de α	τ	ρ_S
Producto	N.A.	0	0
Gaussiana	$-1 \leq \alpha \leq 1$	$\frac{2}{\pi} \text{sen}^{-1}(\alpha)$	$\frac{6}{\pi} \text{sen}^{-1}(\frac{\alpha}{2})$
Clayton	$\alpha > 0$	$\frac{\alpha}{\alpha+2}$	*
Frank	$\alpha \in (-\infty, \infty)$	$1 - \frac{4}{\alpha}[1 - D_1(\alpha)]$	$1 - \frac{12}{\alpha}[D_1(\alpha) - D_2(\alpha)]$
FGM	$-1 \leq \alpha \leq 1$	$\frac{2}{9}\alpha$	$\frac{1}{3}\alpha$

Cuadro 4.1: Algunas familias de cópulas comunes y las relaciones entre α , ρ_S y τ .

Nota: La $D_k(k)$ representa la función $k/x^k \int_0^x \frac{t^k}{e^t-1} dt$, $k = 1, 2$ y FGM representa la cópula Farlie-Gumbel-Morgenstern. El símbolo * indica que la función es complicada [17],[22]. Para más detalle véase el apéndice A.

4.2.4. Rango de las medidas de asociación

En el cuadro 4.2 se muestran los rangos de las medidas ρ_S y τ para diversas familias de cópulas.

Familia	Rango de τ	Rango de ρ_S
FGM	$-\frac{2}{9} \leq \tau \leq \frac{2}{9}$	$-\frac{1}{3} \leq \rho_S \leq \frac{1}{3}$
Gaussiana	$-1 \leq \tau \leq 1$	$-1 \leq \rho_S \leq 1$
Clayton	$0 \leq \tau \leq 1$	$0 \leq \rho_S \leq 1$
Frank	$-1 \leq \tau \leq 1$	$-1 \leq \rho_S \leq 1$

Cuadro 4.2: Rangos de τ y ρ_S de las diversas familias de cópulas.

Observe que en este caso no todos los valores de τ y ρ_S asociadas a las familias de cópulas del cuadro 4.2 cumplen con la primera propiedad, es decir, que no tienen valores entre -1 y +1, como es el caso de la familia FGM (Farlie-Gumbel-Morgenstern), cuyos valores están acotados entre $[-\frac{2}{9}, \frac{2}{9}]$. Esta característica indica un rango limitado de dependencia en la familia en cuestión que restringe su capacidad para modelar. Además, en el caso de la familia Clayton, la medida de asociación toma únicamente valores positivos¹, lo cual indica su incapacidad para identificar dependencias negativas. Por otro lado, la familia Frank provee una fuerte asociación entre las variables ya que es una familia completa [26].

En las figuras 4.7 a 4.10 se muestra el comportamiento de la τ de Kendall para diversos valores del parámetro α en algunas de las familias de cópulas más comunes.

¹La familia Clayton es completa, sin embargo, los valores especificados en el cuadro 4.1 para los valores del parámetro de dependencia la hacen estricta y con ello, incompleta.

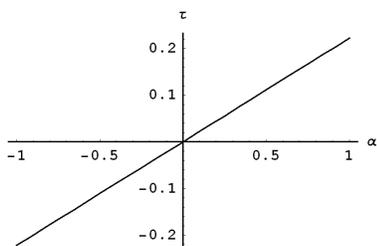


Figura 4.7: Comportamiento de τ con $\alpha \in [-1, 1]$ en la familia FGM.

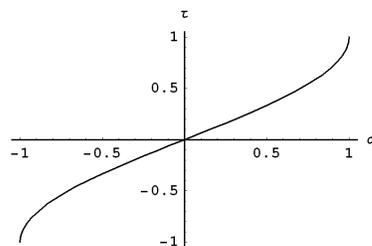


Figura 4.8: Comportamiento de τ con $\alpha \in [-1, 1]$ en la familia Normal.

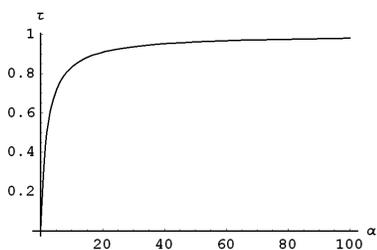


Figura 4.9: Comportamiento de τ con $\alpha \in [0, 100]$ en la familia Clayton.

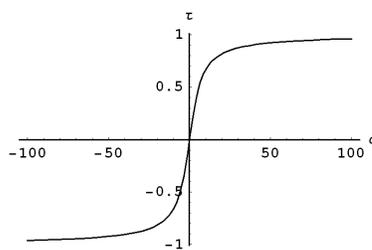


Figura 4.10: Comportamiento de τ con $\alpha \in [-100, 100]$ en la familia Frank.

4.3. Construcción del modelo de probabilidad usando cópulas

Como se ha visto en las secciones anteriores, las propiedades de las diversas familias de cópulas pueden hacerlas ideales para el modelado de distribuciones o para la simulación de datos con una distribución específica. Existen dos técnicas principales para la simulación de datos a partir de cópulas. Una de ellas es el método de *simulación condicional* y la otra es la técnica de *simulación elíptica*.

4.3.1. Simulación condicional

Esta técnica puede ser aplicada a una gran cantidad de cópulas y es común emplearla en cópulas con expresión cerrada como las arquimedianas. En la práctica, el proceso de simulación condicional para dos variables consiste en lo siguiente:

Dado un conjunto de datos inicial P

Paso 1. Calcule τ o ρ_S de P y obtenga el parámetro de dependencia α de la cópula a emplear.

Paso 2. Genere dos valores uniformes (v_1, v_2) de forma aleatoria.

Paso 3. Iguale el primer valor uniforme que se evaluarán en la función de distribución acumulada con el primer valor generado de forma aleatoria ($u_1 = v_1$).

Paso 4. Obtenga el segundo valor uniforme de manera condicional derivando parcialmente la expresión de la cópula empleada con respecto a u_1 e igualándola a v_2 .

Paso 5. Evaluar u_1 y u_2 en las funciones de distribución acumulada marginales para obtener las nuevas observaciones x e y .

En el siguiente ejemplo se muestra la forma de obtener dos nuevas observaciones a partir de un conjunto de datos dado empleando la simulación condicional:

Ejemplo 4.4. Considere el siguiente conjunto de datos de entrada:

X	Y
3.0	4.0
5.0	12.0
-1.0	1.0
1.5	1.3
3.5	-7.0
4.5	-4.0

Si se desean obtener dos nuevas observaciones X_1^{t+1} y Y_1^{t+1} de dicho conjunto empleando una cópula arquimediana de la familia Clayton, se siguen los pasos del procedimiento anterior para simularlos de forma condicional. El primer paso consiste en calcular alguna de las medidas de asociación τ o ρ_s . En este caso, se calculará el valor τ de Kendall por simplicidad, el cual es $\tau=0.2$.

De acuerdo con el cuadro 4.1, τ se encuentra expresado en términos del parámetro de dependencia de la cópula Clayton por $\tau = \frac{\alpha}{\alpha+2}$ y haciendo $\tau=0.2$ se obtiene $\alpha=0.5$.

Enseguida, se generan dos valores uniformes $v_1=0.176$ y $v_2=0.683$ de forma aleatoria y se iguala el primer valor $u_1=v_1=0.176$, lo cual comprende el segundo paso del procedimiento. La forma de $C_\alpha(u_1, u_2)$ en esta familia está dada por:

$$C_\alpha(u_1, u_2) = (u_1^{-\alpha} + u_2^{-\alpha} - 1)^{-1/\alpha},$$

la cual se deriva parcialmente con respecto a u_1 y se iguala con v_2 . Despejando u_2 de esta expresión se llega a:

$$u_2 = (v_1^{-\alpha}(v_2^{-\alpha/(\alpha+1)} - 1) + 1)^{-1/\alpha}.$$

Evaluando los valores de v_1 y v_2 en la expresión anterior se obtiene $u_2=0.895$. Finalmente, para obtener los nuevos datos se evalúan u_1 y u_2 en las funciones de distribución marginales empíricas del conjunto dado y se llega a los dos nuevos valores $X_1^{t+1}=0.202$ y $Y_1^{t+1}=11.701$.

El ejemplo describe la forma de obtener dos nuevas observaciones a partir de un conjunto de datos, empleando la simulación condicional y la cópula de Clayton. Para la regeneración de todo el conjunto inicial se debe repetir el proceso un número N de veces.

4.3.2. Simulación elíptica

Otra técnica para la simulación de datos es el método de *simulación elíptica* el cual es semejante a los métodos empleados para obtener observaciones a partir de distribuciones elípticas como la Normal y la T. Dicha técnica se fundamenta en el siguiente teorema:

Teorema 1 (De la cópula Gaussiana). *Sea \mathbf{Z} un vector de variables distribuidas de forma normal, $H(\cdot)$ su correspondiente función de distribución conjunta y $\Phi(\cdot)$ la función de distribución acumulada estándar normal. Para cualquier conjunto de funciones de distribución marginales F_1, \dots, F_n , las variables*

$$X_1 = F_1^{-1}[\Phi(Z_1)], \dots, X_n = F_n^{-1}[\Phi(Z_n)]$$

tienen una función de distribución conjunta acumulada

$$F_{X_1, \dots, X_n}(x_1, \dots, x_k) = H\{\Phi^{-1}[F_1(x_1)], \dots, \Phi^{-1}[F_n(x_n)]\}.$$

Entonces, en las variables (X_1, \dots, X_n) , se cumple:

$$\begin{aligned}\tau(X_i, X_j) &= \tau(Z_i, Z_j) \\ \rho_s(X_i, X_j) &= \rho_s(Z_i, Z_j)\end{aligned}$$

Este teorema indica que es posible obtener un conjunto de datos que se relacionen de la misma forma que un conjunto de datos de entrada siempre que éstos puedan ser representados por una cópula Normal y que cumplan con las características arriba mencionadas. Con lo cual da la pauta para establecer un sencillo procedimiento de simulación de datos [22]:

Suponga que el vector aleatorio (X_1, \dots, X_n) tiene una distribución normal multivariada con marginales normales $F_{X_1}, \dots, F_{X_n} \sim N(0, 1)$ y una matriz de covarianzas positiva:

$$\Sigma = \begin{pmatrix} 1 & \rho_{12} & \cdots & \rho_{1n} \\ \rho_{21} & 1 & \cdots & \rho_{2n} \\ \vdots & \vdots & & \vdots \\ \rho_{n1} & \rho_{n2} & \cdots & 1 \end{pmatrix}$$

donde $\rho_{ij} = \rho_{ji}$ es el coeficiente de correlación entre X_i y X_j . De forma equivalente a la técnica empleada en el capítulo anterior, se puede obtener una matriz triangular inferior B tal que $\Sigma = BB'$, y mediante la misma técnica allí descrita

se logra obtener un conjunto de datos con cierta correlación deseada. Ahora bien, suponga que se desea obtener un conjunto de datos a partir de la cópula asociada al vector aleatorio dado (X_1, \dots, X_n) y de un coeficiente de correlación por rangos $\tau_{ij} = \tau(X_i, X_j)$ o bien $\rho_S(X_i, X_j)$. Considere el siguiente procedimiento de simulación de Monte Carlo:

Paso 1. Convierta el coeficiente de correlación por rangos dado al coeficiente de correlación de Pearson mediante las equivalencias:

$$\rho_{ij} = \text{sen} \left(\frac{\pi}{2} \tau_{ij} \right) \quad (4.21)$$

$$= 2 \text{sen} \left[\frac{\pi}{6} \rho_S(X_i, X_j) \right]. \quad (4.22)$$

y construya la matriz triangular inferior $B = (b_{ij})$ mediante la ecuación:

$$b_{ij} = \frac{\rho_{ij} - \sum_{s=1}^{j-1} b_{is} b_{js}}{\sqrt{1 - \sum_{s=1}^{j-1} b_{js}^2}}, \quad 1 \leq j \leq i \leq n.$$

Paso 2. Genere un vector columna de variables aleatorias normales independientes $\mathbf{Y} = (Y_1, \dots, Y_n)'$.

Paso 3. Tome la matriz producto de B y \mathbf{Y} : $\mathbf{Z} = (Z_1, \dots, Z_n)' = B\mathbf{Y}$. Obtenga datos uniformes transformados como $U_i = \Phi(Z_i)$, con Φ como la CDF normal estándar y para $i = 1, \dots, n$.

Paso 4. Haga $X_i^{t+1} = F_{X_i}^{-1}(u_i)$ para $i = 1, \dots, n$.

con \mathbf{X}^{t+1} como el nuevo vector distribuido de forma semejante al vector inicial \mathbf{X} .

En el siguiente ejemplo se muestra la forma de aplicar el procedimiento anterior para capturar una distribución conjunta dada y simular datos a partir de ésta. De igual manera, se emplea la técnica del NOPREDA y de las normales independientes para simular nuevos datos a partir de las distribuciones marginales. En cada caso, los resultados son graficados y comparados con la distribución conjunta original para verificar qué tan efectiva es cada técnica para aproximar una distribución dada.

Ejemplo 4.5. Refiérase al ejemplo 4.1 para datos relacionados con el problema.

Para obtener la aproximación de la distribución conjunta $H(x, y)$ empleando cópulas, se generan dos vectores aleatorios u, v con $u \in [0, 1]$ (ya que $x \in [-1, 1]$), y $v \in [0, 1 - e^{-2}]$ (ya que $y \in [0, 2]$). Posteriormente se generan los vectores $F^{[-1]}(u) = 2u - 1$ y $G^{[-1]}(v) = -\ln(1 - v)$ que conformarán la matriz de entrada C . Esta matriz contiene en cada columna una función de distribución marginal cada una de las cuales se pueden asociar a su correspondiente función de distribución conjunta mediante una cópula que realice la correspondencia.

Note que de las expresiones 4.16 y 4.12 se puede ver que para identificar la cópula asociada al par (X, Y) se pueden emplear las medidas de asociación τ de Kendall o ρ_s de Spearman, las cuales contendrán de forma implícita la asociación entre las variables que la cópula le provee.

Así pues, se calcula τ o ρ_s de la matriz de entrada C , y se realiza la conversión al coeficiente de correlación ρ empleando las expresiones indicadas en el paso 1, (4.21 ó 4.22). A partir de este coeficiente, se generan dos nuevos vectores aleatorios que se distribuyen de forma normal multivariada (con matriz de covarianzas Σ integrada por los coeficientes ρ_{ij} calculados en el paso anterior), haciendo uso de la matriz triangular inferior B que cumple $\Sigma = BB'$. Los pasos 1, 2 y 3 describen este proceso. Posteriormente, se obtienen dos vectores unimodales de la evaluación de una función de distribución normal conjunta en los dos vectores obtenidos en los pasos anteriores.

Para generar los nuevos vectores distribuidos como $H(x, y)$, se evalúan las funciones de distribución marginales inversas, en los datos uniformes. Este proceso, similar para el NOPREDA descrito en el capítulo anterior, se realiza mediante una interpolación de los datos de cada uno de los vectores de distribución marginal $F(x)$ y $G(y)$.

En la figura 4.11 se muestra la función de distribución conjunta $H(x, y)$ obtenida de sustituir los nuevos datos en la función 4.3 empleando 500 datos en cada dimensión.

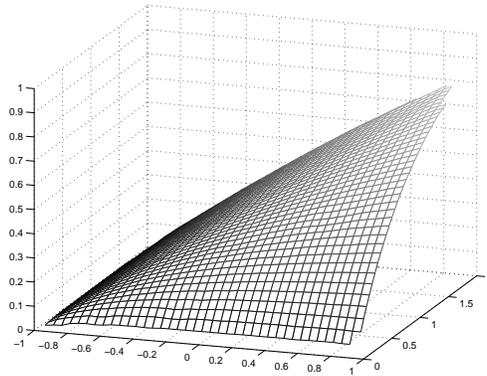


Figura 4.11: Función de distribución conjunta $H(x, y)$ obtenida mediante una cópula normal.

Al comparar la figura 4.1 con 4.11 se puede observar que este procedimiento logra capturar casi a la perfección la distribución de un conjunto de datos que no necesariamente se encuentran distribuidos de forma normal.

En comparación, un proceso similar a éste se puede efectuar con normales independientes y con la técnica del NOPREDA vista en el capítulo anterior para capturar la correlación entre los datos de entrada. En las figuras 4.12 y 4.13 se

pueden apreciar los resultados obtenidos.

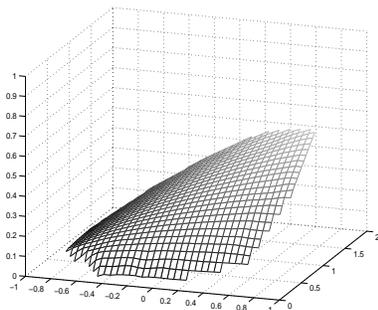


Figura 4.12: Función de distribución conjunta $H(x, y)$ obtenida mediante normales independientes.

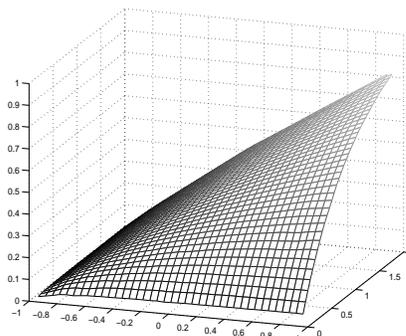


Figura 4.13: Función de distribución conjunta $H(x, y)$ obtenida mediante la técnica del NOPREDA.

Como se puede apreciar, en ambos casos las distribución se aproxima a la original, sin embargo, en el caso de la aproximación mediante normales independientes el parecido es menor que en el caso del NOPREDA o en el de la cópula normal. En el siguiente ejemplo se muestra que en efecto, la dependencia entre las variables está pobremente representada mediante la normales independientes, a diferencia del NOPREDA o mediante cópulas.

Ejemplo 4.6. *Para mostrar experimentalmente la eficacia de los métodos de simulación presentados en el ejemplo anterior, se realizará la siguiente prueba: primeramente se genera un conjunto de datos con cierta distribución y se le calcula su matriz de correlación, enseguida se simularán datos a partir de dicha matriz de correlación utilizando la técnica del NOPREDA, con una cópula normal y con normales independientes; finalmente se obtendrán las matrices de correlación de los datos de todas estas técnicas y se compararán con la matriz original. Si estos métodos capturan de manera satisfactoria las dependencias del conjunto inicial, entonces se obtendrá una matriz de correlación semejante a la original.*

Con lo cual, para probar la hipótesis, se genera un conjunto de datos aleatorios normalmente distribuidos con vector de medias $\mu = (0, 0)$ y matriz de covarianzas $\Sigma = \begin{pmatrix} 1 & 0.7 \\ 0.7 & 1 \end{pmatrix}$. Posteriormente, se calcula su coeficiente de correlación ρ_1 . En segunda instancia, a partir de esta correlación conocida, se simulan datos empleando la técnica del NOPREDA vista en el capítulo anterior y de los datos así obtenidos se calcula el coeficiente de correlación ρ_2 . Luego, mediante normales independientes, se obtiene una nueva población de datos.

Recuerde que la técnica de las normales independientes supone que cada instancia i del vector columna j de la matriz de entrada C puede regenerarse como

4.3. Construcción del modelo de probabilidad usando cópulas

$C_{ij}^{t+1} = \mu_j + \sigma_j^2 z$ donde μ_j es la media de la j -ésima columna de C , σ_j^2 es la varianza de la j -ésima columna de C y z es un número aleatorio con distribución normal $(0,1)$.

A los datos obtenidos con normales independientes se les calcula su coeficiente de correlación ρ_3 . Finalmente, empleando cópulas normales, como se describió en el ejemplo anterior, se regenera la población y su matriz de correlaciones ρ_4 se compara con las matrices anteriores.

$$\rho_1 = \begin{pmatrix} 1.0000 & 0.7313 \\ 0.7313 & 1.0000 \end{pmatrix} \quad \rho_2 = \begin{pmatrix} 1.0000 & 0.7145 \\ 0.7145 & 1.0000 \end{pmatrix}$$

$$\rho_3 = \begin{pmatrix} 1.0000 & 0.0044 \\ 0.0044 & 1.0000 \end{pmatrix} \quad \rho_4 = \begin{pmatrix} 1.0000 & 0.7407 \\ 0.7407 & 1.0000 \end{pmatrix}$$

Observe que la correlación entre los datos regenerados mediante el NOPREDA o mediante cópulas se aproxima muy bien a la matriz de correlaciones original. El caso de las normales independientes, sin embargo, se muestra muy alejado de los valores deseables. En las figuras 4.14 a 4.17 se ha graficado la primera columna de C^{t+1} contra la segunda columna de la misma matriz para cada uno de los casos antes descritos.

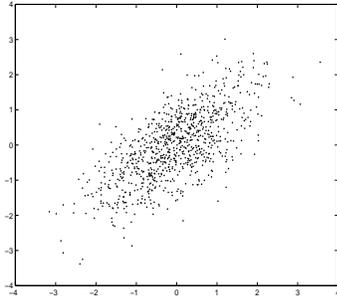


Figura 4.14: Gráfica C_1^t contra C_2^t de los datos originales.

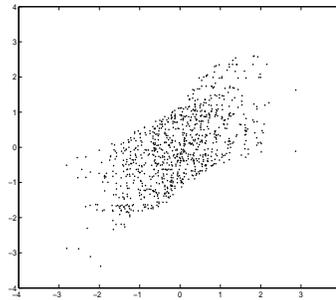


Figura 4.15: Gráfica C_1^{t+1} contra C_2^{t+1} de los datos simulados con NOPREDA.

Observe en la figura 4.16 los datos simulados mediante normales independientes. Como era de esperarse las relaciones entre ambos vectores es pobre y la gráfica resultante indica independencia entre las nuevas columnas, característica indeseable para una técnica de modelación. Por otro lado, el NOPREDA y las cópulas muestran una gran semejanza con los datos originales.

Como se ha observado en los ejemplos anteriores, la técnica de las cópulas permite capturar las dependencias de los datos originales para generar una nueva población con esta distribución. Ésta es precisamente la base del algoritmo de

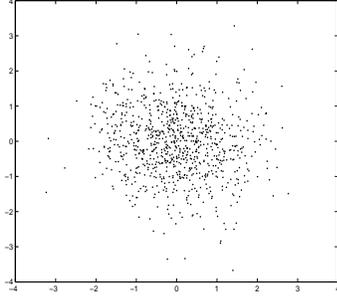


Figura 4.16: Gráfica C_1^{t+1} contra C_2^{t+1} de los datos simulados mediante normales independientes.

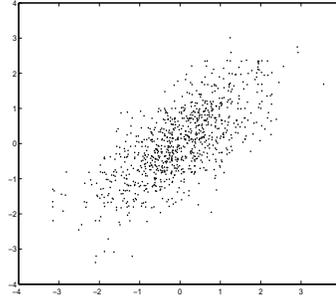


Figura 4.17: Gráfica C_1^{t+1} contra C_2^{t+1} de los datos simulados con una cópula normal.

estimación de distribución con cópulas, misma que se muestra y se detalla en la sección siguiente.

4.4. Algoritmo de Estimación de Distribución con Cópulas: COPULEDA

El estudio de las cópulas para la simulación de datos con cierta distribución ha abierto la posibilidad de emplear dicha herramienta en los algoritmos de estimación de distribución para inducir la correlación de los datos de la muestra hacia la nueva generación. Esta es la idea principal del COPULEDA, algoritmo que se indica en el algoritmo 4 y se enuncia con detenimiento a continuación.

Algoritmo 4 COPULEDA

- 1: $t=0$
 - 2: Generar N individuos de forma aleatoria y guardarlos en $P(t)$.
 - 3: **while** no se cumpla la condición de paro **do**
 - 4: Evaluar los individuos de $P(t)$. Seleccionar $M \leq N$ y colocarlos en S .
 - 5: Calcular τ o ρ_S de S , convertirlo a ρ empleando 4.21 ó 4.22 y guardarlo en Σ .
 - 6: Generar $N \times n$ valores Z distribuidos $\mathcal{N}(\mathbf{0}, \Sigma)$.
 - 7: **for** $j = 1$ hasta n **do**
 - 8: **for** $i = 1$ hasta N **do**
 - 9: Obtener los valores uniformes transformados $u_{i,j} = \Phi(z_{i,j})$.
 - 10: Generar $P(t+1)_{i,j} = F_{X_j}^{-1}(u_{i,j})$ con F_{X_j} como la función de distribución acumulada empírica de la variable j .
 - 11: **end for**
 - 12: **end for**
 - 13: $t=t+1$
 - 14: **end while**
-

4.4. Algoritmo de Estimación de Distribución con Cópulas: COPULEDA

El algoritmo de estimación de distribución con cópulas se describe en el algoritmo 4 y los 5 primeros pasos coinciden con el NOPREDA. En el paso 5 comienza el proceso de simulación elíptica descrito en la sección anterior y consta del cálculo de la medida de asociación τ o ρ_S de la muestra S . Esta medida posee las características grupales de la muestra y es convertida a su correspondiente valor de correlación ρ para ser empleada en la cópula gaussiana y con ella en el proceso de simulación tipo Monte Carlo.

Es importante hacer notar aquí que las equivalencias entre las medidas de asociación y el coeficiente de correlación no fueron implementadas para las pruebas del COPULEDA, sino que se dejó la aproximación con τ o ρ_S debido a la facilidad con la que la matriz de muestra se volvía singular en poblaciones pequeñas.

La generación de datos multivariados a partir de una matriz de correlación dada, ha sido explicada con anterioridad en las secciones 3.2 y 4.3. Ambas técnicas describen un proceso en el que se multiplica una matriz triangular inferior (que puede resultar del proceso de factorización de Cholesky) por una matriz de vectores aleatorios, lo más decorrelacionados posible. En el caso del NOPREDA, esta matriz resultante dispone el orden de una nueva matriz de valores, los cuales se desean asociar de manera semejante a los datos de la muestra. En este caso, estos datos que ya poseen la correlación deseada, sirven como *seleccionadores* de los datos uniformes a partir de los cuales se hará el remuestreo. Esto se realiza evaluando en la función de distribución estándar los valores aleatorios generados en el paso anterior, observe la figura 4.18.

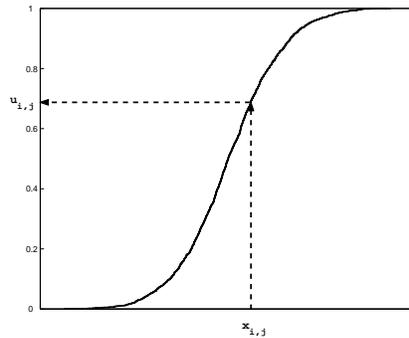


Figura 4.18: Obtención de los datos uniformes a partir de la distribución acumulada estándar.

Pese a que no la función de distribución acumulada estándar no se puede expresar de forma cerrada, existen diversas formas de realizar algunas aproximaciones para valores dados de la variable aleatoria. Para más detalles, revise [27].

Los datos uniformes así obtenidos estarán correlacionados de acuerdo con los datos de la muestra y ya pueden emplearse para generar la nueva población.

El proceso de regeneración es igual que en el algoritmo del NOPREDA, pero en este caso ya no se requiere realizar reindexación alguna y es aquí donde finaliza una generación del COPULEDA.

En el siguiente ejemplo se muestra el resultado de aplicar el procedimiento del COPULEDA para la regeneración de una población inicial dada.

Ejemplo 4.7. *Si la población inicial es como se muestra en la figura 3.4 y se aplica el procedimiento del COPULEDA para la regeneración de los individuos, se obtiene el resultado que aparece en la figura 4.19.*

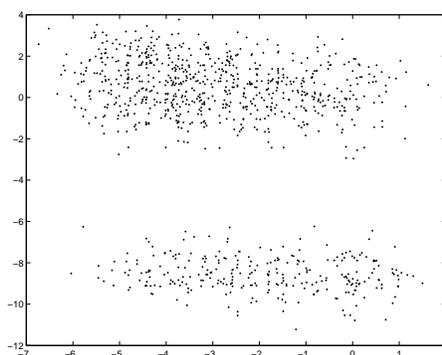


Figura 4.19: Población regenerada mediante COPULEDA $\rho_S = -0.3100$

Como se puede apreciar, el resultado es semejante al obtenido por NOPREDA sobre la misma población inicial. Al igual que en este caso, el coeficiente de correlación por rangos se aproxima de buena manera al coeficiente inicial de $\rho_S = -0.3143$.

4.5. Ajuste de la varianza

Uno de los principales problemas de los algoritmos de estimación de distribución es la gran velocidad de convergencia que ocasiona que muchos de los problemas de optimización se queden estancados en mínimos locales. El ajuste de la varianza es un tema que ha sido estudiado e implementado en este tipo de algoritmos para tratar de evitar la convergencia prematura así como los mínimos locales [28]. En estos algoritmos, el ajuste de la varianza se reduce al producto de la matriz de covarianzas por un factor de escalamiento que permite abrir o cerrar los valores de cada variable en un valor fijo, de acuerdo con cierto parámetro indicativo. Este parámetro, basado en relaciones entre los valores de las variables y la función de aptitud, marca la pauta para el ajuste de la varianza y se efectúa repetidamente en todo el proceso del algoritmo hasta cumplir con el criterio de paro.

En el caso del NOPREDA y el COPULEDA, la tendencia a aproximar la distribución de la cual provienen los datos ocasiona un aumento considerable en la velocidad de convergencia que es posible frenar mediante un ajuste de varianza similar al comentado para otros algoritmos.

En la figura 4.20 se muestra el efecto del aumento o disminución de la varianza en la función de densidad de una variable X determinada, distribuida de forma gaussiana. Observe que en estos casos, la media de los datos se mantiene constante en cero y no se pierde la distribución original. Estas características son deseables en cualquier técnica de ajuste de varianza, ya que en toda búsqueda aleatoria es primordial no perder *del todo* al mejor individuo de la población y explorar con cierta libertad en sus prometedoras cercanías.

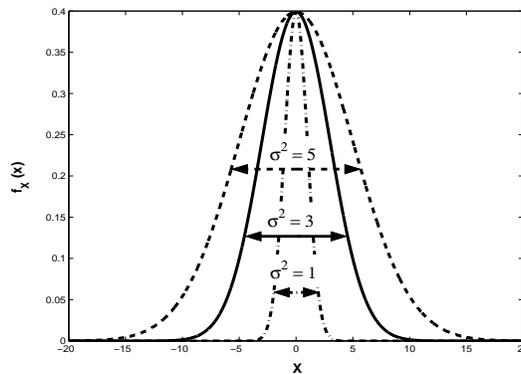


Figura 4.20: Ajuste de la varianza en la densidad de la variable aleatoria X .

La forma de ajustar la varianza que se propone en el presente trabajo, de igual forma, intenta mantener constante la media de los datos así como la distribución original de la muestra. A manera de ilustración, suponga, que los datos de una de las variables de la muestra se encuentran *colapsados* en un mínimo local y debido a las características de regeneración de los algoritmos descritos en el presente capítulo y en el anterior, es casi imposible obtener nuevas instancias que diversifiquen la población. Con esto, es deseable que exista una técnica cuya finalidad sea separar los valores de su centro de masa, con cierta holgura y reinicie la búsqueda en un ambiente más propicio.

La técnica empleada consta de reajustar los valores de la variable aleatoria en cuestión a partir de un factor de escalamiento cuyo valor cambia de acuerdo con el número de datos en la población y la media de la variable aleatoria.

4.5.1. Transformación lineal

Primero, considérese el caso simple en el que el factor de escalamiento es constante. Sea entonces

$$x_{i,j}^{t+1} = (x_{i,j}^t - \mu_j)\sqrt{F} + \mu_j \quad (4.23)$$

la relación de ajuste de varianza, donde $x_{i,j}$ es el i -ésimo valor que toma la variable aleatoria j , μ_j es la media de la j -ésima variable y F es el factor de escalamiento.

La ecuación 4.23 indica que a cada uno de los datos de la variable aleatoria a ajustar se le resta su correspondiente media y se multiplica por el factor de escalamiento con la finalidad de aumentar o reducir la varianza en un factor de F y finalmente colocar nuevamente los datos sobre su media original. Los valores que puede tomar F están entre 0 y 1 si se desea disminuir la varianza y mayores que uno si lo que se quiere es aumentar la varianza. Observe el siguiente ejemplo:

Ejemplo 4.8. Sea X una variable aleatoria uniforme $(-100,100)$ y μ su correspondiente media. Tras la ejecución de un programa de computadora se generaron diez muestras de la variable $X = (67.125, -79.115, 61.171, 98.510, -64.353, 3.689, 88.054, 94.292, -31.525, 91.001)$ y se calcularon $\mu_X = 32.885$ y $\sigma_X^2 = 4830.187$.

Suponga que se desea aumentar la varianza de X al doble, entonces se aplica la ecuación 4.23 con $F = 2$ para cada uno de los elementos de X . Con ello se obtiene el vector $X' = (81.308, -125.507, 72.887, 125.692, -104.630, -8.403, 110.906, 119.728, -58.205, 115.073)$. Es fácil verificar que este vector posee una media de $\mu_{X'} = 32.885$ y que tiene una varianza de $\sigma_{X'}^2 = 9660.374$, la cual es exactamente el doble de la calculada en X .

En el ejemplo 4.8 se pudo ver que a partir de un vector de valores aleatorios y de la expresión 4.23 es posible ajustar la varianza para generar un segundo vector con la misma media. Además de estas características, esta transformación lineal mantiene la distribución de los datos originales, siempre que éstos provengan de una distribución normal.

4.5.2. Transformación no lineal

En los ejemplos anteriores la transformación de los datos se realizó de forma constante en cada uno de ellos. Ahora bien, suponga que el factor de escalamiento F no toma un único valor constante, si no que se va aumentando o disminuyendo de acuerdo con el número de datos que se encuentren en la muestra, de tal forma que los datos más cercanos al cúmulo central se desvían de él sólo una pequeña proporción, mientras que los más alejados se alejan aún más, permitiendo una búsqueda menos lineal.

En este trabajo de investigación el proceso de ajuste de varianza está basado precisamente en esta idea. La ecuación que se resuelve en cada iteración de reajuste es la siguiente:

$$F_{i+1,j} = \eta F_{i,j} \quad (4.24)$$

donde η es una variable de decaimiento o incremento por la cual se multiplica el factor $F_{i,j}$ cada vez que se ajusta el dato i de la variable j . Entonces, siguiendo esta idea, para ajustar un valor $x_{i,j}$ determinado, se debe conocer la media de las instancias de la variable a ajustar y a partir de ésta comenzar a separar los datos que se encuentren hacia su izquierda y hacia su derecha en una proporción, la cual aumenta o disminuye conforme los datos se alejan de la media, observe la figura 4.21.

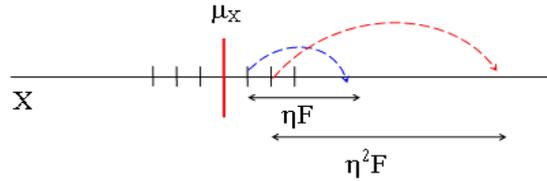


Figura 4.21: Ajuste de varianza de una conjunto de datos.

A manera de ilustración, considérese el siguiente ejemplo:

Ejemplo 4.9. *Suponga que se desean ajustar los valores de la variable aleatoria X_j de tal forma que los valores transformados tengan una varianza del doble de la original, entonces $F = 2$. Los datos han sido ordenados y cinco de ellos se encuentran hacia la derecha de la media. Se desea ajustar la varianza empleando la ecuación 4.23 pero que a cada instancia de la variable X_j el factor de escalamiento se reajuste como indica la expresión 4.24, de tal forma que cuando se terminen de ajustar los cinco datos la varianza final haya llegado a $F_{5,j} = 2$, esto es:*

$$\begin{aligned} F_{5,j} &= \eta F_{4,j} = 2 \\ &= \eta \cdot \eta \cdot F_{3,j} = \eta \cdot \eta \cdot \eta \cdot F_{2,j} \\ &= \eta^4 F_{1,j} = 2 \end{aligned}$$

Si se resuelve esta expresión, se llega al valor de η para que dicha igualdad se cumpla:

$$\eta = \left(\frac{2}{F_{1,j}} \right)^{\frac{1}{4}}.$$

Observe que aunque se ha fijado la máxima varianza final de los datos, la mínima $F_{1,j}$ no ha quedado establecida. Tómese entonces, como mínimo valor de abertura $F_{1,j} = 1$ y calcúlese el valor de η :

$$\eta = (2)^{\frac{1}{4}} \approx 1.1892$$

Tomar $F_{1,j} = 1$ indica que el primer dato a la derecha de la media se transformará en sí mismo luego del ajuste de la varianza, sin embargo, $F_{1,j}$ no se encuentra limitado a este valor y depende del lector elegir un valor adecuado para esta transformación inicial. En el caso del presente trabajo, se ha elegido comenzar con un valor inicial de η calculado como en el ejemplo anterior y terminar en $\eta^{n_{\text{máx}}}$ con $n_{\text{máx}}$ como el máximo número de datos a ajustar por la derecha, con lo cual la expresión general es

$$\eta = (F_{\text{máx}})^{\frac{1}{n_{\text{máx}}}}, \quad (4.25)$$

donde $F_{\text{máx}}$ es el máximo valor de varianza deseada. En el caso de los datos hacia la izquierda de la media se procede de forma similar a la anterior, únicamente tomando a consideración que en este caso los datos serán el reflejo de los que se encuentren del lado derecho de la media.

Este procedimiento, pese a que fija la varianza máxima que tomarán los datos, no determina la varianza global que tendrán éstos, como lo era en el caso de una F fija. Sin embargo, como se verá en el siguiente ejemplo, el resultado es muy próximo al máximo establecido.

Ejemplo 4.10. *Se desea ajustar el vector $X = (-5, -4, -3, -2, -1, 0, 1, 2, 3, 4, 5)$ de tal forma que la máxima variación sea de $F_{\text{máx}} = 2$, entonces, se comienza ordenando los datos que se encuentren por la derecha y por la izquierda de la media ($\mu=0$), de tal forma que se tengan dos nuevos vectores $X_1 = (-5, -4, -3, -2, -1)$ y $X_2 = (1, 2, 3, 4, 5)$. Enseguida se calcula η para los valores de X_2 siguiendo la ecuación 4.25:*

$$\eta = 2^{\frac{1}{5}} \approx 1.1487$$

Entonces se comenzará con $F_1 = \eta = 1.1487$ y se emplea este valor para obtener el primer dato transformando, empleando la ecuación 4.23:

$$\begin{aligned} y_1 &= (x_1 - \mu)\sqrt{F_1} + \mu \\ &= (1)\sqrt{1.1487} \\ &\approx 1.0717. \end{aligned}$$

Luego, para la transformación del siguiente dato, se actualiza el valor de F_2 :

$$\begin{aligned} F_2 &= \eta F_1 \\ &= (1.1487)^2 \\ &\approx 1.3195, \end{aligned}$$

nuevamente este valor es empleado para calcular el siguiente valor de Y . Este proceso se repite hasta que se haya aplicado la transformación a todos los valores de X_2 . En el caso de los valores de X_1 , como se encuentran ordenados de forma inversa a X_2 , se utiliza la expresión:

$$\eta = (F_{\text{máx}})^{\frac{-1}{n_{\text{máx}}}}, \quad (4.26)$$

la cual puede ser fácilmente deducida luego de un proceso semejante al visto en el ejemplo 4.9. Al terminar de transformar todos los datos se obtiene $Y = (-7.071, -5.278, -3.693, -2.297, -1.071, 0, 1.071, 2.297, 3.693, 5.278, 7.071)$, incluyendo, por supuesto, la media. Se puede observar que en este caso la media es constante para los dos vectores y que las varianzas son $\sigma_X^2 = 11$, $\sigma_Y^2 = 19.585$ que es un poco menos del doble de la original.

En la figura 4.22 se han graficado los datos originales y los transformados. Como se puede apreciar, los datos nuevos siguen una transformación no lineal que contrasta con los originales o con los obtenidos por el procedimiento de factor fijo visto en la subsección anterior.

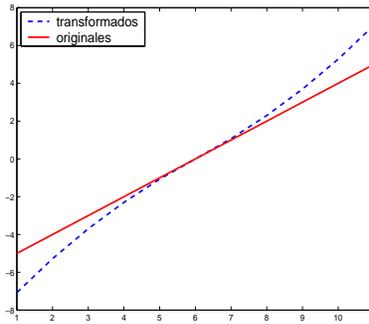


Figura 4.22: Gráfica de los datos originales y los datos con transformación $F_{\text{máx}} = 2$.

En la siguiente sección se describe una técnica sencilla para aplicar el ajuste de la varianza tanto en el COPULEDA como en el NOPREDA.

4.5.3. Tiempo de ajuste

Como se ha visto en las subsecciones anteriores, reajustar la varianza de las variables para transformar los datos de una población puede ser útil para controlar la velocidad de convergencia de un algoritmo de búsqueda aleatoria como es el caso de los algoritmos de estimación de distribución. Sin embargo, determinar cuándo aplicar estos ajustes de varianza es un problema casi tan importante como el método de búsqueda en sí.

En este caso, lo deseable sería que cada ajuste de varianza se efectuara de manera automática dependiendo de la condición de los individuos en las generaciones, sin embargo, en la actualidad no existe una forma universal de autoajustar variables cuya distribución se desconoce y hacerlo de forma eficiente.

En este trabajo de investigación se propone el siguiente procedimiento de ajuste, el cual describe una heurística sencilla para el incremento o decremento de la varianza.

```
1: t=0
2: if t mod k = 0 then
3:   Observar el desempeño del nuevo mejor individuo de la población y
   guardarlo en  $g$ .
4:   if  $g > \delta$  then
5:     Aumentar la varianza si antes se ha aumentado.
6:     Disminuir la varianza si antes se ha disminuido.
7:   else
8:     Aumentar la varianza si antes se ha disminuido.
9:     Disminuir la varianza si antes se ha aumentado.
10:  end if
11: end if
12: t=t+1
```

Los valores de k y δ deben ser determinados por el lector, así como el factor de ajuste de varianza.

La calidad del desempeño del nuevo mejor individuo se puede apreciar como una razón entre la aptitud del individuo elite actual y el t módulo k -ésimo anterior. Un problema de este tipo de pruebas es que revela una escala no normalizada lo cual dificulta la elección del factor δ .

4.6. Conclusiones del capítulo 4

En el presente capítulo se introdujo un nuevo algoritmo de estimación de distribución que modela la distribución de los datos originales a partir de los datos de la muestra utilizando una función de cópula gaussiana. Se vio que una cópula es una función que permite realizar la relación entre la función de distribución conjunta y sus marginales de bajo orden y que ello implica que la función posea de forma implícita las relaciones entre estas marginales. Se mostraron los diferentes tipos de cópulas y sus principales propiedades así como algunas técnicas para generar cópulas. Se definieron las medidas de asociación de Kendall y Spearman en términos de cópulas, las propiedades deseables de una medida de asociación y las relaciones universales entre ellas para cualquier familia de cópulas. Asimismo, se mostró la forma de relacionar las medidas de asociación de acuerdo con el parámetro de dependencia y los valores que toma de acuerdo con la familia de cópulas empleada. Se mostró la técnica para la simulación de valores aleatorios utilizando cópulas, misma que fue la base teórica del algoritmo de estimación de distribución con cópulas (COPULEDA), el cual fue presentado utilizando una cópula gaussiana. Posteriormente se habló sobre algunas técnicas para la reducción o aumento de varianza y su utilización en los algoritmos de búsqueda estocástica.

4.6. Conclusiones del capítulo 4

Capítulo 5

Experimentos y resultados

En el presente capítulo se muestran los experimentos realizados para la prueba del NOPREDA y el COPULEDA, así como los resultados arrojados por estos algoritmos en varias dimensiones. Algunos comentarios sobre las pruebas y los resultados obtenidos son mostrados al final del capítulo.

5.1. Conjuntos experimentales

Las pruebas realizadas se dividen en tres conjuntos de acuerdo con el tipo de problema a resolver. El primero abarca diez problemas de minimización, cada uno de ellos es un problema unimodal convexo y es probado en varias dimensiones. Para estos diez problemas, la región de búsqueda es $[-10,5]$. El segundo conjunto está conformado por siete problemas multimodales, seis de ellos de minimización y uno de maximización, en todos ellos las regiones de búsqueda del óptimo son diferentes, de tal forma que los tres primeros comparten la región de búsqueda del primer conjunto, y los cuatro restantes cumplen la región de búsqueda descrita en la literatura ([29], [30]). Todos ellos se prueban en varias dimensiones. Finalmente, el tercer conjunto está compuesto por tres funciones de maximización tanto unimodales como multimodales, las cuales se distinguen de los conjuntos anteriores porque carecen de pendientes que dirijan la población hacia alguna zona específica de la región de búsqueda. Para dos de ellas, el espacio de búsqueda es $[0,10]$, mientras que para la otra es de $[-0.16,0.16]$.

El criterio de paro en todos estos problemas fue que se alcanzara el óptimo con una precisión de $1E-10$, o bien que se agotara el número máximo de iteraciones. Todos los pormenores de los resultados obtenidos se describen la siguiente sección.

A continuación se muestran los conjuntos experimentales.

Conjunto experimental A

Las funciones que se muestran enseguida se probaron en diversas dimensiones y se caracterizan por ser unimodales y convexas.

1. $f_{A,1} = \sum_{i=1}^n x_i^2$
2. $f_{A,2} = \sum_{i=1}^n 10^{6 \frac{i-1}{n-1}} x_i^2.$
3. $f_{A,3} = x_1^2 + \sum_{i=2}^n 10^6 x_i^2.$
4. $f_{A,4} = 10^6 x_1^2 + \sum_{i=2}^n x_i^2.$
5. $f_{A,5} = x_1^2 + \sum_{i=2}^{n-1} 10^4 x_i^2 + 10^8 x_n^2.$
6. $f_{A,6} = \sum_{i=1}^{\lfloor n/2 \rfloor} 10^6 x_i^2 + \sum_{i=\lfloor n/2 \rfloor}^n x_i^2.$
7. $f_{A,7} = \sum_{i=1}^n |x_i|^{2+10 \frac{i-1}{n-1}}.$
8. $f_{A,8} = \sum_{i=1}^{n-1} [100 \times (x_i^2 - x_{i+1})^2 + (x_i - 1)^2].$
9. $f_{A,9} = -x_1 + 100 \sum_{i=2}^n x_i^2.$
10. $f_{A,10} = -x_1 + 100 \sqrt{\sum_{i=2}^n x_i^2}.$

Las figuras 5.1 y 5.2 muestran dos de las funciones más representativas del conjunto experimental A .

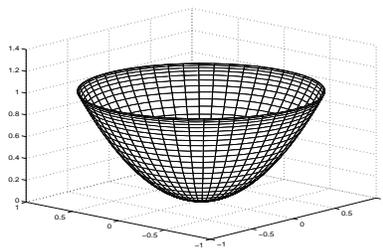


Figura 5.1: Función $f_{A,1}$

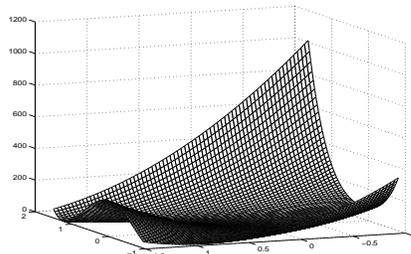


Figura 5.2: Función $f_{A,8}$

Conjunto experimental B

Las siguientes funciones son multimodales y se probaron en diversas dimensiones, la región de búsqueda del óptimo es la misma que para el conjunto experimental A.

$$1. f_{B,1} = -20 \exp \left(-0.2 \sqrt{\frac{1}{n} \sum_{i=1}^n x_i^2} \right) - \exp \left(\frac{1}{n} \sum_{i=1}^n \cos(2\pi x_i) \right) + 20 + \exp(1).$$

$$2. f_{B,2} = \frac{1}{4000} \sum_{i=1}^n x_i^2 - \prod_{i=1}^n \cos \left(\frac{x_i}{\sqrt{i}} \right).$$

$$3. f_{B,3} = \sum_{i=1}^n [x_i^2 - 10 \cos(2\pi x_i) + 10].$$

$$4. f_{B,4} = \frac{\pi}{n} \left\{ 10 \operatorname{sen}^2(\pi y_1) + \sum_{i=1}^{n-1} (y_i - 1)^2 [1 + 10 \operatorname{sen}^2(\pi y_{i+1})] + (y_n - 1)^2 \right\} + \sum_{i=1}^n u(x_i, 10, 100, 4)$$

donde

$$y_i = 1 + \frac{1}{4}(x_i + 1)$$

y

$$u(x_i, a, k, m) = \begin{cases} k(x_i - a)^m & x_i > a \\ 0 & -a \leq x_i \leq a \\ k(-x_i - a)^m & x_i < -a \end{cases}$$

$$5. f_{B,5} = \frac{1}{10} \left\{ \operatorname{sen}^2(3\pi x_1) + \sum_{i=1}^{n-1} (x_i - 1)^2 [1 + \operatorname{sen}^2(3\pi x_{i+1})] \right\} + \frac{1}{10} (x_n - 1)^2 [1 + \operatorname{sen}^2(2\pi x_n)] + \sum_{i=1}^n u(x_i, 5, 100, 4).$$

$$6. f_{B,6} = \sum_{i=1}^5 [(\vec{x} - \vec{a}_i)(\vec{x} - \vec{a}_i)^T + \vec{c}_i]^{-1}$$

i	$a_{i,j}$	$j = 1, \dots, n$	c_i
1	2	2 2 ... 2	0.1
2	4	4 4 ... 4	0.2
3	8	8 8 ... 8	0.2
4	6	6 6 ... 6	0.4
5	3	7 3 ... 7	0.4

$$7. f_{B,7} = \sum_{i=1}^n \left[-x_1 \text{sen}(\sqrt{|x_i|}) \right].$$

Observe en las figuras 5.3 a 5.7 algunas de las funciones más representativas del conjunto experimental B .

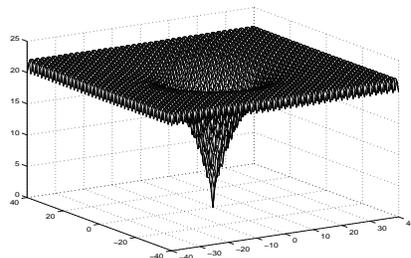


Figura 5.3: Función $f_{B,1}$

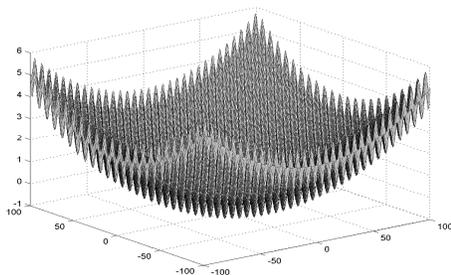


Figura 5.4: Función $f_{B,2}$

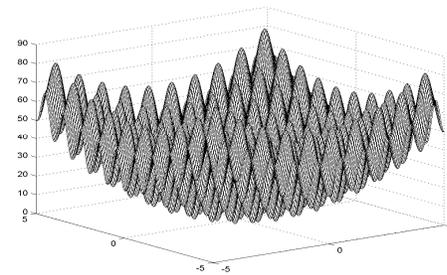


Figura 5.5: Función $f_{B,3}$

Conjunto experimental C

Las siguientes funciones tienen la particularidad de estar formadas por uno o varios máximos locales, sobre una región de búsqueda que tiene ligeras o nulas pendientes, con lo cual muchos de los programas de optimización como el EDA, se pierden en el proceso de búsqueda del máximo.

$$1. f_{C,1} = \frac{1}{10^{-5} + \sum_{i=1}^n |y_i|}$$

donde

$$\begin{aligned} y_1 &= x_1 \\ y_i &= x_i + y_{i-1}, \quad i \geq 2. \end{aligned}$$

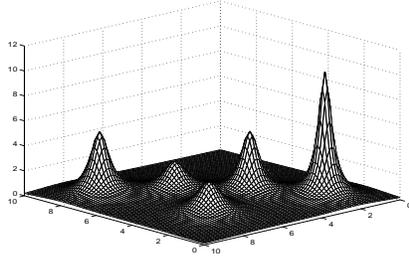


Figura 5.6: Función $f_{B,6}$

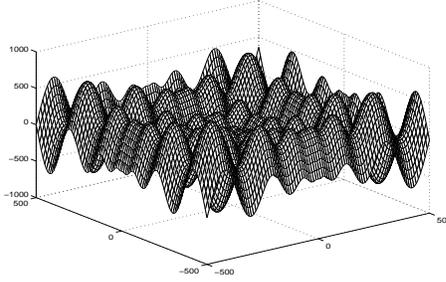


Figura 5.7: Función $f_{B,7}$

$$2. f_{C,2} = \sum_{i=1}^2 \alpha_i \mathcal{N}(\vec{x}, \mu_i, \Sigma_i)$$

donde $\mathcal{N}(\vec{x}, \mu_i, \Sigma_i)$ es una distribución normal multivariada la cual tiene un vector n -dimensional de medias μ y una matriz de covarianzas Σ de $n \times n$. Además, $\alpha_1 = 1000$, $\alpha_2 = 900$, $\mu_1 = (-10, \dots, -10)$ y $\mu_2 = (10, \dots, 10)$, mientras que $\Sigma_i (i = 1, 2)$ son matrices diagonales cuyos elementos de la diagonal son iguales a 1.

$$3. f_{C,3} = \sum_{i=1}^3 \alpha_i \mathcal{N}(\vec{x}, \mu_i, \Sigma_i).$$

con los mismos parámetros que $f_{C,2}$ pero además con $\alpha_3 = 500$ y $\mu_3 = (0, \dots, 0)$.

En las figuras 5.8 y 5.9 se muestran algunas de las funciones representativas del conjunto experimental C .

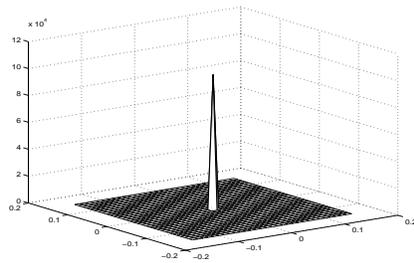


Figura 5.8: Función $f_{C,1}$

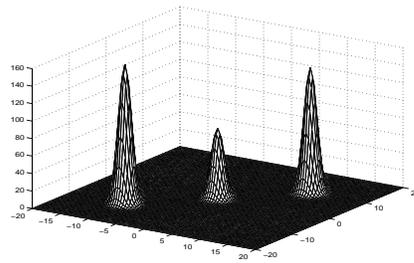


Figura 5.9: Función $f_{C,3}$

5.2. Resultados

Los óptimos de cada función son mostrados en el cuadro 5.1.

5.2. Resultados

Conjunto A			Conjunto B			Conjunto C		
$F(x)$	Tipo	Óptimo	$F(x)$	Tipo	Óptimo	$F(X)$	Tipo	Óptimo
$f_{A,1}$	Mín.	0	$f_{B,1}$	Mín.	0	$f_{C,1}$	Máx.	10^5
$f_{A,2}$	Mín.	0	$f_{B,2}$	Mín.	0	$f_{C,2}$	Máx.	10.1053
$f_{A,3}$	Mín.	0	$f_{B,3}$	Mín.	0	$f_{C,3}$	Máx.	10.1053
$f_{A,4}$	Mín.	0	$f_{B,4}$	Mín.	0			
$f_{A,5}$	Mín.	0	$f_{B,5}$	Mín.	0			
$f_{A,6}$	Mín.	0	$f_{B,6}$	Máx.	10.1033 [†]			
$f_{A,7}$	Mín.	0	$f_{B,7}$	Mín.	-2.0949E+03 [‡]			
$f_{A,8}$	Mín.	0						
$f_{A,9}$	Mín.	$-\infty$						
$f_{A,10}$	Mín.	$-\infty$						

Cuadro 5.1: Las funciones y sus óptimos.[†]La función $f_{B,6}$ tiene sus óptimos en 10.1033 y 10.0134, para $n = 4$ y 30, respectivamente. [‡]La función $f_{B,7}$ tiene sus óptimos en -2.0949E+03, -4.1898E+03, -6.2847E+03 y -8.3796E+03 para $n = 5, 10, 15$ y 20, respectivamente. Los óptimos de $f_{A,9}$ y $f_{A,10}$ se han fijado en -1.00E+10, por simplicidad.

5.2.1. NOPREDA sin ajuste de varianza

En los cuadros 5.2 a 5.12 se muestran los resultados al ejecutar el NOPREDA sin ajuste de varianza en 5, 10, 15 y 20 dimensiones con un máximo número de evaluaciones de 60000 y una población de 30 individuos. El criterio de paro empleado fue que el algoritmo convergiera a la solución con una precisión de $1E - 10$ o bien que alcanzara el número máximo de evaluaciones.

Conjunto A

Los cuadros 5.2 a 5.5 muestran los resultados arrojados por el NOPREDA en 5, 10, 15 y 20 dimensiones sin ajuste de varianza para el primer conjunto experimental. Los campos mostrados indican el mejor individuo de 10 corridas independientes, el peor, la media, la mediana y la desviación estándar de estas diez corridas así como el promedio de evaluaciones.

$F(x)$	Mejor	Peor	Media	Mediana	D.E.	Eval.
$f_{A,1}$	3.7419E-12	7.7924E-08	1.1240E-08	4.3781E-09	1.7549E-08	58903
$f_{A,2}$	2.5236E-08	3.6324E-03	3.1831E-04	3.9308E-05	7.8978E-04	60000
$f_{A,3}$	3.2225E-06	2.3156E-02	5.1803E-03	1.4233E-03	6.9364E-03	60000
$f_{A,4}$	8.3333E-10	2.1368E-02	1.0159E-03	5.1216E-06	3.9521E-03	60000
$f_{A,5}$	1.0018E-06	5.1129E-01	3.1741E-02	1.4557E-03	9.8625E-02	60000
$f_{A,6}$	1.0113E-05	6.7037E-02	6.9050E-03	1.2915E-03	1.3729E-02	60000
$f_{A,7}$	9.1705E-14	6.3697E-09	3.6072E-10	2.5923E-11	1.1922E-09	36207
$f_{A,8}$	7.0006E-01	1.8153E+00	1.6108E+00	1.8040E+00	3.3455E-01	60000
$f_{A,9}$	-1.0000E+10	-1.0000E+10	-1.0000E+10	-1.0000E+10	2.7888E-06	53713
$f_{A,10}$	-1.0000E+10	-1.0000E+10	-1.0000E+10	-1.0000E+10	4.7114E-03	60000

Cuadro 5.2: Resultados de NOPREDA sin ajuste de varianza en 5 dimensiones

5.2. Resultados

$F(x)$	Mejor	Peor	Media	Mediana	D.E.	Eval.
$f_{A,1}$	1.2261E-11	2.0910E-08	2.3664E-09	8.5841E-10	4.0399E-09	59676
$f_{A,2}$	1.1475E-06	2.1248E-02	3.0251E-03	5.7033E-04	5.5533E-03	60000
$f_{A,3}$	1.8403E-05	1.9470E-01	1.3840E-02	3.0247E-03	3.8966E-02	60000
$f_{A,4}$	5.3517E-09	6.1708E-01	2.5031E-02	2.3311E-05	1.1252E-01	60000
$f_{A,5}$	4.8096E-07	4.2728E+01	2.0938E+00	3.0712E-03	8.1906E+00	60000
$f_{A,6}$	3.9111E-04	1.8856E-01	4.0173E-02	1.1417E-02	5.4681E-02	60000
$f_{A,7}$	9.3094E-15	4.4289E-09	2.1741E-10	4.9935E-11	7.9953E-10	37463
$f_{A,8}$	1.1511E+00	1.8262E+00	1.7893E+00	1.8136E+00	1.2090E-01	60000
$f_{A,9}$	-1.0000E+10	-1.0000E+10	-1.0000E+10	-1.0000E+10	3.1079E-06	58449
$f_{A,10}$	-1.0000E+10	-1.0000E+10	-1.0000E+10	-1.0000E+10	8.9998E-04	60000

Cuadro 5.3: Resultados de NOPREDA sin ajuste de varianza en 10 dimensiones.

$F(x)$	Mejor	Peor	Media	Mediana	D.E.	Eval.
$f_{A,1}$	3.5793E-12	1.1518E-08	4.4242E-10	6.1953E-11	2.0922E-09	47810
$f_{A,2}$	4.1330E-08	1.9570E-02	1.8312E-03	1.2030E-05	4.3491E-03	60000
$f_{A,3}$	9.5583E-08	8.1945E-03	3.2503E-04	9.2760E-06	1.4943E-03	60000
$f_{A,4}$	3.6021E-09	5.8997E-02	8.7830E-03	4.1765E-04	1.7578E-02	60000
$f_{A,5}$	8.4430E-08	2.8311E+02	9.5435E+00	7.7494E-04	5.1670E+01	60000
$f_{A,6}$	1.1332E-05	7.4375E-01	4.4171E-02	1.5325E-02	1.3614E-01	60000
$f_{A,7}$	9.8837E-15	9.9989E-11	4.6235E-11	3.4025E-11	3.5205E-11	28775
$f_{A,8}$	5.5475E-02	1.8279E+00	1.6181E+00	1.8163E+00	4.3922E-01	60000
$f_{A,9}$	-1.0000E+10	-1.0000E+10	-1.0000E+10	-1.0000E+10	3.0877E-06	60000
$f_{A,10}$	-1.0000E+10	-1.0000E+10	-1.0000E+10	-1.0000E+10	2.3913E-03	60000

Cuadro 5.4: Resultados de NOPREDA sin ajuste de varianza en 15 dimensiones

$F(x)$	Mejor	Peor	Media	Mediana	D.E.	Eval.
$f_{A,1}$	7.2222E-12	4.0866E-08	1.7415E-09	9.8156E-11	7.4202E-09	56694
$f_{A,2}$	3.1130E-06	4.2623E-03	7.7504E-04	2.0758E-04	1.2346E-03	60000
$f_{A,3}$	1.9325E-06	5.5966E-04	1.1106E-04	3.2824E-05	1.5908E-04	60000
$f_{A,4}$	2.8550E-09	2.6091E-01	1.5241E-02	2.8537E-04	5.1136E-02	60000
$f_{A,5}$	5.1385E-06	1.4425E+01	7.2168E-01	4.0192E-03	2.8178E+00	60000
$f_{A,6}$	2.1790E-04	2.8852E-01	4.1268E-02	1.0419E-02	7.0214E-02	60000
$f_{A,7}$	2.1980E-12	9.9869E-11	5.7105E-11	6.5098E-11	3.3831E-11	25807
$f_{A,8}$	5.7768E-03	1.8358E+00	1.6085E+00	1.8270E+00	5.2586E-01	60000
$f_{A,9}$	-1.0000E+10	-1.0000E+10	-1.0000E+10	-1.0000E+10	5.6113E-06	60000
$f_{A,10}$	-1.0000E+10	-1.0000E+10	-1.0000E+10	-1.0000E+10	5.8812E-03	60000

Cuadro 5.5: Resultados de NOPREDA sin ajuste de varianza en 20 dimensiones

Conjunto B

En los cuadros 5.6 a 5.9 se muestran los resultados arrojados por el NOPREDA en 5, 10, 15 y 20 dimensiones para el conjunto de prueba B , con excepción del problema $f_{B,6}$ el cual fue probado únicamente en dimensiones 4 y 30. Los resultados para este problema se muestran al final de esta subsección.

5.2. Resultados

$F(x)$	Mejor	Peor	Media	Mediana	D.E.	Eval.
$f_{B,1}$	2.5318E-06	4.5791E-04	1.6207E-04	1.2165E-04	1.3739E-04	60000
$f_{B,2}$	3.7962E-09	3.4477E-02	1.5271E-02	1.2320E-02	8.2651E-03	60000
$f_{B,3}$	2.4502E-08	3.4768E-04	1.5202E-05	1.1804E-06	6.3108E-05	60000
$f_{B,4}$	7.5735E-12	9.4750E-07	4.7953E-08	2.9110E-10	1.7641E-07	57535
$f_{B,5}$	1.0883E-12	1.0167E-05	3.4495E-07	1.4314E-10	1.8553E-06	58013
$f_{B,7}$	2.0949E+03	2.0949E+03	2.0949E+03	2.0949E+03	2.7386E-04	20299

Cuadro 5.6: Resultados de NOPREDA sin ajuste de varianza en 5 dimensiones.

$F(x)$	Mejor	Peor	Media	Mediana	D.E.	Eval.
$f_{B,1}$	7.0840E-06	5.0907E-04	1.0763E-04	7.0859E-05	1.1655E-04	60000
$f_{B,2}$	7.6580E-11	4.4282E-02	1.2720E-02	1.2316E-02	9.6703E-03	59767
$f_{B,3}$	1.6927E-08	2.9786E-05	2.6669E-06	4.9449E-07	5.8209E-06	60000
$f_{B,4}$	8.6824E-12	5.3772E-09	4.6840E-10	8.0902E-11	1.2348E-09	55595
$f_{B,5}$	3.4875E-12	1.6974E-08	9.9662E-10	9.7898E-11	3.1800E-09	55906
$f_{B,7}$	4.1898E+03	4.1898E+03	4.1898E+03	4.1898E+03	1.0916E-04	26859

Cuadro 5.7: Resultados de NOPREDA sin ajuste de varianza en 10 dimensiones

$F(x)$	Mejor	Peor	Media	Mediana	D.E.	Eval.
$f_{B,1}$	7.5211E-08	1.7361E-05	3.8251E-06	2.8136E-06	3.4150E-06	60000
$f_{B,2}$	1.3754E-11	3.1957E-02	1.0178E-02	9.8572E-03	8.0868E-03	56295
$f_{B,3}$	5.7326E-11	2.1806E-07	3.7078E-08	1.4157E-08	5.7663E-08	59926
$f_{B,4}$	1.6763E-11	4.1673E-08	1.5226E-09	9.3925E-11	7.5855E-09	52302
$f_{B,5}$	6.0344E-12	5.8168E-10	9.5248E-11	7.4288E-11	1.1212E-10	50689
$f_{B,7}$	6.2847E+03	6.2847E+03	6.2847E+03	6.2847E+03	2.0795E-04	21485

Cuadro 5.8: Resultados de NOPREDA sin ajuste de varianza en 15 dimensiones

$F(x)$	Mejor	Peor	Media	Mediana	D.E.	Eval.
$f_{B,1}$	2.4086E-06	3.1063E-05	1.1527E-05	9.0585E-06	7.6539E-06	60000
$f_{B,2}$	3.2734E-11	3.4447E-02	7.7977E-03	7.3960E-03	9.5009E-03	56211
$f_{B,3}$	1.2237E-09	2.6936E-06	1.6984E-07	3.5972E-08	4.9694E-07	60000
$f_{B,4}$	5.9990E-11	2.8219E-09	5.1632E-10	1.8446E-10	6.9772E-10	58041
$f_{B,5}$	6.1001E-11	2.6028E-08	1.5613E-09	4.2707E-10	4.7286E-09	57927
$f_{B,7}$	8.3796E+03	8.3796E+03	8.3796E+03	8.3796E+03	5.2461E-05	29980

Cuadro 5.9: Resultados de NOPREDA sin ajuste de varianza en 20 dimensiones

Dim.	Mejor	Peor	Media	Mediana	D.E.	Eval.
4	1.0103E+01	2.6369E+00	4.7400E+00	5.1025E+00	2.4437E+00	55495
30	5.0138E+00	5.0135E+00	5.0138E+00	5.0138E+00	8.4145E-05	60000

Cuadro 5.10: Resultados de NOPREDA sin ajuste de varianza para $f_{B,6}$.

Conjunto C

En los cuadros 5.11 y 5.12 se muestran los resultados arrojados por el NOPREDA en 5, 10, 15 y 20 dimensiones para la función $f_{C,1}$ y en 5 dimensiones

5.2. Resultados

para las restantes del conjunto C .

$F(x)$	Mejor	Peor	Media	Mediana	D.E.	Eval.
$f_{C,1}$	9.6377E+04	2.2851E+03	3.1800E+04	2.0354E+04	2.9155E+04	60000
$f_{C,2}$	1.0105E+01	0.0000E+00	5.9954E+00	9.0944E+00	4.6565E+00	57923
$f_{C,3}$	1.0105E+01	0.0000E+00	5.7895E+00	7.0735E+00	4.2411E+00	58283

Cuadro 5.11: Resultados de NOPREDA sin ajuste de varianza en 5 dimensiones.

Dim.	Mejor	Peor	Media	Mediana	D.E.	Eval.
10	7.6143E+04	1.2882E+03	2.2584E+04	1.6702E+04	1.9682E+04	60000
15	9.4669E+04	1.3050E+03	1.8833E+04	1.0411E+04	2.5087E+04	60000
20	6.4226E+04	2.7156E+02	1.0629E+04	6.7664E+03	1.4028E+04	60000

Cuadro 5.12: Resultados de $f_{C,1}$ con NOPREDA sin ajuste de varianza.

Las dimensiones y regiones de búsqueda de $f_{B,6}$, $f_{C,1}$, $f_{C,2}$ y $f_{C,3}$ se fijaron para coincidir con las propuestas por Lu y Yao [30]. Observe la dificultad de los problemas propuestos, los cuales fueron diseñados especialmente para destacar en algoritmos que emplean aglomeraciones en su búsqueda, razón por la cual el NOPREDA se desempeña de manera poco eficiente.

5.2.2. NOPREDA con ajuste de varianza

Al igual que en la sección anterior, los tres conjuntos experimentales fueron probados en el NOPREDA pero realizando un ajuste en la varianza, si es que era necesaria, cada dos generaciones. A la nueva población se le agregaron tres mutaciones del mejor individuo en un factor de $l_p=0.98$ (esto es que el número que se le agrega al valor elite es un aleatorio $(0, l_p)$) y se trabajó con 30 individuos en la población.

Conjunto A

En los cuadros 5.13 a 5.16 se muestran los resultados obtenidos por el NOPREDA en 5, 10, 15 y 20 dimensiones para el conjunto experimental A.

$F(x)$	Mejor	Peor	Media	Mediana	D.E.	Eval.
$f_{A,1}$	6.3954E-12	9.9806E-11	7.2600E-11	8.2051E-11	2.5204E-11	20583
$f_{A,2}$	1.1668E-12	9.9809E-11	5.5315E-11	5.9037E-11	3.4422E-11	35706
$f_{A,3}$	2.9619E-12	9.9525E-11	5.9062E-11	6.4005E-11	3.3107E-11	34712
$f_{A,4}$	5.9628E-14	2.2331E-10	6.6070E-11	6.3351E-11	4.2150E-11	30801
$f_{A,5}$	1.8628E-14	7.1249E-08	2.4168E-09	4.2300E-11	1.3000E-08	42435
$f_{A,6}$	3.8507E-13	2.6405E-07	8.8451E-09	4.7014E-11	4.8200E-08	41235
$f_{A,7}$	7.3366E-14	9.8416E-11	5.1086E-11	4.8778E-11	3.4753E-11	9956
$f_{A,8}$	1.1646E-02	1.7225E+00	6.6265E-01	5.6714E-01	5.9520E-01	60000
$f_{A,9}$	-1.0000E+10	-1.0000E+10	-1.0000E+10	-1.0000E+10	0.0000E+00	20438
$f_{A,10}$	-1.0000E+10	-1.0000E+10	-1.0000E+10	-1.0000E+10	1.0625E-06	58351

Cuadro 5.13: Resultados de NOPREDA con ajuste de varianza en 5 dimensiones

5.2. Resultados

$F(x)$	Mejor	Peor	Media	Mediana	D.E.	Eval.
$f_{A,1}$	4.8573E-11	9.9816E-11	8.7532E-11	9.0472E-11	1.2943E-11	23274
$f_{A,2}$	1.9117E-12	9.9843E-11	7.5263E-11	8.3922E-11	2.4496E-11	35832
$f_{A,3}$	1.7555E-11	9.9982E-11	7.1593E-11	8.3562E-11	2.5205E-11	39708
$f_{A,4}$	1.2895E-11	4.5451E-04	4.7818E-05	8.5671E-11	1.2218E-04	46557
$f_{A,5}$	1.3367E-13	4.5732E-04	1.5244E-05	6.0597E-11	8.3496E-05	46500
$f_{A,6}$	5.8578E-12	9.9871E-11	6.9582E-11	7.9269E-11	3.0509E-11	36587
$f_{A,7}$	4.3956E-12	9.9631E-11	6.7716E-11	7.3751E-11	2.8350E-11	16031
$f_{A,8}$	9.9230E-02	1.7226E+00	1.1900E+00	1.5681E+00	5.6946E-01	60000
$f_{A,9}$	-1.0000E+10	-1.0000E+10	-1.0000E+10	-1.0000E+10	1.6986E-06	55548
$f_{A,10}$	-1.0000E+10	-1.0000E+10	-1.0000E+10	-1.0000E+10	3.8474E-06	60000

Cuadro 5.14: Resultados de NOPREDA con ajuste de varianza en 10 dimensiones

$F(x)$	Mejor	Peor	Media	Mediana	D.E.	Eval.
$f_{A,1}$	4.8784E-11	9.9984E-11	8.3676E-11	8.7282E-11	1.4403E-11	32088
$f_{A,2}$	3.2261E-11	1.1511E-09	1.7925E-10	9.8682E-11	2.2044E-10	58291
$f_{A,3}$	1.0261E-09	2.2335E-07	3.6354E-08	1.5314E-08	5.1897E-08	60000
$f_{A,4}$	3.9404E-11	2.2327E-02	8.8097E-04	1.2217E-10	4.1020E-03	53512
$f_{A,5}$	3.2703E-10	1.1325E-03	3.8321E-05	9.8865E-09	2.0667E-04	60000
$f_{A,6}$	2.5450E-11	3.2732E-09	2.6467E-10	9.5258E-11	6.2796E-10	56534
$f_{A,7}$	1.1460E-11	9.8678E-11	7.8251E-11	8.5615E-11	2.1395E-11	18506
$f_{A,8}$	1.4631E-01	1.7825E+00	1.4005E+00	1.6635E+00	4.4288E-01	60000
$f_{A,9}$	-1.0000E+10	-1.0000E+10	-1.0000E+10	-1.0000E+10	1.8060E-06	60000
$f_{A,10}$	-1.0000E+10	-1.0000E+10	-1.0000E+10	-1.0000E+10	2.1094E-05	60000

Cuadro 5.15: Resultados de NOPREDA con ajuste de varianza en 15 dimensiones

$F(x)$	Mejor	Peor	Media	Mediana	D.E.	Eval.
$f_{A,1}$	7.7486E-11	4.5055E-10	1.7293E-10	1.3546E-10	9.4789E-11	57995
$f_{A,2}$	5.3402E-08	8.9834E-07	3.3269E-07	2.6165E-07	2.1270E-07	60000
$f_{A,3}$	1.1812E-05	3.6205E-04	8.7429E-05	6.0203E-05	8.2097E-05	60000
$f_{A,4}$	6.5054E-10	2.6277E-02	1.2946E-03	1.1767E-07	5.2231E-03	60000
$f_{A,5}$	1.5058E-06	4.3831E-03	3.3056E-04	1.8825E-05	8.8024E-04	60000
$f_{A,6}$	3.3639E-08	1.6446E-06	3.2391E-07	1.7755E-07	3.9530E-07	60000
$f_{A,7}$	7.0634E-12	9.9905E-11	7.6321E-11	8.2572E-11	2.3777E-11	21549
$f_{A,8}$	2.1960E-01	1.7993E+00	1.2995E+00	1.5236E+00	5.1642E-01	60000
$f_{A,9}$	-1.0000E+10	-1.0000E+10	-1.0000E+10	-1.0000E+10	3.9440E-06	60000
$f_{A,10}$	-1.0000E+10	-1.0000E+10	-1.0000E+10	-1.0000E+10	6.0787E-04	60000

Cuadro 5.16: Resultados de NOPREDA con ajuste de varianza en 20 dimensiones

Conjunto B

En los cuadros 5.17 a 5.20 se muestran los resultados obtenidos por el NOPREDA realizando ajuste de varianza en 5, 10, 15 y 20 dimensiones para algunos de los problemas del conjunto experimental B y al igual que en el caso anterior, los resultados de $f_{B,6}$ en 4 y 30 dimensiones son mostrados al final de esta subsección en el cuadro 5.21.

5.2. Resultados

$F(x)$	Mejor	Peor	Media	Mediana	D.E.	Eval.
$f_{B,1}$	7.9731E-12	9.9164E-11	7.3088E-11	8.2958E-11	2.7380E-11	44037
$f_{B,2}$	6.7417E-11	3.2013E-02	1.4778E-02	1.3546E-02	9.0993E-03	54490
$f_{B,3}$	7.4180E-12	9.9632E-11	6.1208E-11	7.0983E-11	3.1539E-11	25220
$f_{B,4}$	1.7678E-12	9.7925E-11	6.3242E-11	6.2950E-11	2.5587E-11	20662
$f_{B,5}$	1.5107E-11	9.9292E-11	6.6961E-11	7.3849E-11	2.6601E-11	20782
$f_{B,7}$	2.0949E+03	2.0949E+03	2.0949E+03	2.0949E+03	2.9380E-04	7380

Cuadro 5.17: Resultados de NOPREDA con ajuste de varianza en 5 dimensiones.

$F(x)$	Mejor	Peor	Media	Mediana	D.E.	Eval.
$f_{B,1}$	4.2470E-11	6.9538E-10	1.9666E-10	1.4374E-10	1.6341E-10	58324
$f_{B,2}$	7.6648E-11	5.9008E-02	1.5101E-02	9.8572E-03	1.3937E-02	56148
$f_{B,3}$	1.2860E-11	9.9987E-11	8.4248E-11	9.2072E-11	2.0067E-11	29481
$f_{B,4}$	1.6356E-11	9.9995E-11	8.0752E-11	9.1769E-11	2.3757E-11	23683
$f_{B,5}$	2.6586E-11	9.9854E-11	8.0736E-11	8.8508E-11	1.9800E-11	23345
$f_{B,7}$	4.1898E+03	4.1898E+03	4.1898E+03	4.1898E+03	5.1164E-05	8877

Cuadro 5.18: Resultados de NOPREDA con ajuste de varianza en 10 dimensiones.

$F(x)$	Mejor	Peor	Media	Mediana	D.E.	Eval.
$f_{B,1}$	7.4235E-08	9.7138E-07	2.6878E-07	2.2292E-07	1.9707E-07	60000
$f_{B,2}$	2.3699E-11	1.7236E-02	3.9423E-03	9.7384E-11	5.3578E-03	42047
$f_{B,3}$	3.1604E-11	9.9987E-11	7.5947E-11	7.7236E-11	2.2840E-11	47120
$f_{B,4}$	3.6225E-11	9.9530E-11	8.4077E-11	9.1938E-11	1.8991E-11	34931
$f_{B,5}$	1.9082E-11	9.9772E-11	7.7086E-11	8.4481E-11	2.3974E-11	35911
$f_{B,7}$	6.2847E+03	6.2847E+03	6.2847E+03	6.2847E+03	1.3516E-04	10726

Cuadro 5.19: Resultados de NOPREDA con ajuste de varianza en 15 dimensiones.

$F(x)$	Mejor	Peor	Media	Mediana	D.E.	Eval.
$f_{B,1}$	3.7143E-06	1.6767E-05	1.0791E-05	1.0186E-05	3.6163E-06	60000
$f_{B,2}$	5.8784E-11	5.8784E-11	5.9140E-03	7.3960E-03	5.3973E-03	56612
$f_{B,3}$	2.0966E-08	1.8726E-05	7.4540E-07	6.9151E-08	3.3994E-06	60000
$f_{B,4}$	8.8394E-10	1.0814E-08	4.0262E-09	3.1479E-09	2.8549E-09	60000
$f_{B,5}$	1.1787E-09	1.4507E-08	5.2841E-09	4.1511E-09	3.7807E-09	60000
$f_{B,7}$	8.3796E+03	8.3796E+03	8.3796E+03	8.3796E+03	1.7184E-05	17723

Cuadro 5.20: Resultados de NOPREDA con ajuste de varianza en 20 dimensiones.

Dim.	Mejor	Peor	Media	Mediana	D.E.	Eval.
4	1.0103E+01	2.6369E+00	4.7377E+00	5.1025E+00	2.4487E+00	53437
30	5.0138E+00	2.5253E+00	4.8479E+00	5.0138E+00	6.3134E-01	60000

Cuadro 5.21: Resultados de NOPREDA con ajuste de varianza en $f_{B,6}$.

Conjunto C

Al igual que en el caso sin ajuste de varianza, las funciones $f_{C,2}$ y $f_{C,3}$ fueron probadas únicamente en 5 dimensiones, cuyos resultados se muestran en conjunto con $f_{C,1}$ en el cuadro 5.11, mientras que el cuadro 5.12 muestra los

5.2. Resultados

resultados de $f_{C,1}$ en las dimensiones restantes.

$F(x)$	Mejor	Peor	Media	Mediana	D.E.	Eval.
$f_{C,1}$	1.0000E+05	4.3759E+04	9.5321E+04	9.9948E+04	1.3572E+04	60000
$f_{C,2}$	1.0105E+01	0.0000E+00	8.4835E+00	1.0105E+01	3.4127E+00	35133
$f_{C,3}$	1.0105E+01	0.0000E+00	7.8484E+00	9.0947E+00	2.5685E+00	48376

Cuadro 5.22: Resultados de NOPREDA con ajuste de varianza en 5 dimensiones.

Dim.	Mejor	Peor	Media	Mediana	D.E.	Eval.
10	1.0000E+05	6.3718E+04	9.3157E+04	9.9514E+04	1.0649E+04	60000
15	1.0000E+05	3.0107E+04	9.5417E+04	9.9998E+04	1.5723E+04	60000
20	9.9999E+04	8.4968E+04	9.8366E+04	9.9916E+04	3.5708E+03	60000

Cuadro 5.23: Resultados de $f_{C,1}$ con NOPREDA y ajuste de varianza.

Observe que aún en este caso, los resultados arrojados por el NOPREDA para $f_{C,2}$ y $f_{C,3}$ no son tan buenos como se esperaba. Sin embargo, otras corridas realizadas con este algoritmo pero con un número mayor de individuos, arroja mejores resultados pese a que el número de evaluaciones se mantiene constante.

5.2.3. COPULEDA sin ajuste de varianza

En el caso del COPULEDA, se probaron los mismos conjuntos experimentales que en el caso del NOPREDA, con las mismas dimensiones y espacios de búsqueda. Además, al igual que en el caso anterior, los resultados se indican con y sin ajuste en la varianza. En los cuadros siguientes se muestran los resultados para cada uno de los conjuntos experimentales.

Conjunto A

En los cuadros 5.24 a 5.27 se muestran los resultados obtenidos al ejecutar el algoritmo del COPULEDA sin ajuste de varianza para el primer conjunto experimental en 5, 10, 15 y 20 dimensiones.

$F(x)$	Mejor	Peor	Media	Mediana	D.E.	Eval.
$f_{A,1}$	3.5726E-13	9.8360E-011	4.6772E-11	4.2362E-11	3.4973E-11	33550
$f_{A,2}$	9.1697E-14	3.1574E-006	2.1860E-07	8.7915E-11	7.5804E-07	47352
$f_{A,3}$	1.5452E-12	1.4333E-004	5.3004E-06	1.4009E-10	2.6097E-05	55051
$f_{A,4}$	2.5025E-13	7.7025E-002	2.5932E-03	9.2674E-11	1.4059E-02	43565
$f_{A,5}$	8.1870E-12	4.9149E-003	2.1963E-04	1.3618E-08	9.1403E-04	56669
$f_{A,6}$	3.1470E-12	3.5315E-004	1.2159E-05	8.4736E-08	6.4409E-05	56340
$f_{A,7}$	1.1683E-12	9.4164E-011	4.3103E-11	4.2640E-11	2.8383E-11	15208
$f_{A,8}$	1.4638E-04	2.7213E+000	1.0691E+00	5.9206E-01	1.0031E+00	60000
$f_{A,9}$	-1.0000E+10	-1.0000E+010	-1.0000E+10	-1.0000E+10	1.1200E-06	38089
$f_{A,10}$	-1.0000E+10	-1.0000E+010	-1.0000E+10	-1.0000E+10	3.0262E-06	60000

Cuadro 5.24: Resultados de COPULEDA sin ajuste de varianza en 5 dimensiones.

5.2. Resultados

$F(x)$	Mejor	Peor	Media	Mediana	D.E.	Eval.
$f_{A,1}$	8.4288E-11	3.1645E-08	1.8156E-09	6.0045E-10	5.7002E-09	59991
$f_{A,2}$	3.9263E-07	8.0154E-04	4.2393E-05	4.6129E-06	1.4786E-04	60000
$f_{A,3}$	1.4239E-06	3.4813E-03	4.2138E-04	2.1646E-04	7.7627E-04	60000
$f_{A,4}$	3.2400E-11	4.4112E-02	1.5031E-03	1.9493E-07	8.0486E-03	59917
$f_{A,5}$	1.4440E-07	5.6058E-01	2.4733E-02	8.4299E-05	1.0505E-01	60000
$f_{A,6}$	9.3844E-07	2.1651E-02	1.2182E-03	5.0728E-05	4.0574E-03	60000
$f_{A,7}$	6.9458E-12	9.9325E-11	6.5016E-11	7.4506E-11	3.0915E-11	22045
$f_{A,8}$	1.0823E-03	7.0648E+00	1.7165E+00	6.7329E-01	2.2540E+00	60000
$f_{A,9}$	-1.0000E+10	-1.0000E+10	-1.0000E+10	-1.0000E+10	3.6466E-06	60000
$f_{A,10}$	-1.0000E+10	-1.0000E+10	-1.0000E+10	-1.0000E+10	4.5524E-03	60000

Cuadro 5.25: Resultados de COPULEDA sin ajuste de varianza en 10 dimensiones.

$F(x)$	Mejor	Peor	Media	Mediana	D.E.	Eval.
$f_{A,1}$	7.9387E-13	9.9399E-11	5.4340E-11	5.0744E-11	3.2242E-11	42835
$f_{A,2}$	7.1727E-09	8.3788E-03	4.8410E-04	9.6905E-06	1.6305E-03	60000
$f_{A,3}$	3.1697E-09	6.1890E-05	3.7777E-06	7.4197E-07	1.1295E-05	60000
$f_{A,4}$	5.3727E-11	2.6959E-02	1.8802E-03	1.6546E-06	5.7437E-03	58870
$f_{A,5}$	1.3518E-08	9.2660E-01	6.7440E-02	1.0509E-04	1.9025E-01	60000
$f_{A,6}$	4.4745E-05	4.5174E-02	6.6271E-03	1.1617E-03	1.1096E-02	60000
$f_{A,7}$	5.0393E-12	9.9401E-11	5.3561E-11	5.5316E-11	3.1133E-11	24078
$f_{A,8}$	4.5925E-03	1.8185E+00	7.9553E-01	5.3772E-01	7.4317E-01	60000
$f_{A,9}$	-1.0000E+10	-1.0000E+10	-1.0000E+10	-1.0000E+10	2.6027E-06	60000
$f_{A,10}$	-1.0000E+10	-1.0000E+10	-1.0000E+10	-1.0000E+10	1.1146E-04	60000

Cuadro 5.26: Resultados de COPULEDA sin ajuste de varianza en 15 dimensiones.

$F(x)$	Mejor	Peor	Media	Mediana	D.E.	Eval.
$f_{A,1}$	2.5315E-11	1.3492E-09	1.6090E-10	9.5989E-11	2.4683E-10	54298
$f_{A,2}$	4.5075E-07	3.3378E-03	2.7643E-04	1.7497E-05	7.2298E-04	60000
$f_{A,3}$	3.1991E-07	2.3938E-03	2.1808E-04	2.8617E-05	5.5192E-04	60000
$f_{A,4}$	6.5654E-10	8.6589E-02	6.0047E-03	6.0544E-05	1.6423E-02	60000
$f_{A,5}$	2.2208E-07	3.6950E-01	1.8555E-02	4.1475E-04	6.7666E-02	60000
$f_{A,6}$	5.2151E-06	5.8783E-02	3.5006E-03	4.7512E-04	1.0794E-02	60000
$f_{A,7}$	1.0336E-13	9.9746E-11	6.3107E-11	6.7444E-11	3.0022E-11	24055
$f_{A,8}$	1.7159E-03	1.8269E+00	8.4362E-01	5.7536E-01	7.2313E-01	60000
$f_{A,9}$	-1.0000E+10	-1.0000E+10	-1.0000E+10	-1.0000E+10	3.7650E-06	60000
$f_{A,10}$	-1.0000E+10	-1.0000E+10	-1.0000E+10	-1.0000E+10	1.1345E-03	60000

Cuadro 5.27: Resultados de COPULEDA sin ajuste de varianza en 20 dimensiones.

Conjunto B

En los cuadros 5.28 a 5.31 se muestran los resultados obtenidos al ejecutar el COPULEDA sin ajuste de varianza en 5, 10, 15 y 20 dimensiones para la mayoría de los problemas del conjunto B . Los resultados del problema $f_{B,6}$ en 4 y 30 dimensiones se muestran en el cuadro 5.32.

5.2. Resultados

$F(x)$	Mejor	Peor	Media	Mediana	D.E.	Eval.
$f_{B,1}$	9.9935E-10	3.5942E-06	3.1658E-07	5.7175E-08	7.4615E-07	60000
$f_{B,2}$	4.6205E-11	3.2013E-02	1.2516E-02	1.2316E-02	6.8350E-03	57778
$f_{B,3}$	1.3216E-12	8.3163E-09	4.0826E-10	5.1539E-11	1.5566E-09	43205
$f_{B,4}$	9.9898E-13	2.9170E-08	1.1128E-09	7.3334E-11	5.3066E-09	50826
$f_{B,5}$	1.1013E-12	5.1531E-08	3.2342E-09	6.2954E-11	1.0111E-08	52108
$f_{B,7}$	2.0949E+03	2.0949E+03	2.0949E+03	2.0949E+03	2.4851E-11	60000

Cuadro 5.28: Resultados de COPULEDA sin ajuste de varianza en 5 dimensiones

$F(x)$	Mejor	Peor	Media	Mediana	D.E.	Eval.
$f_{B,1}$	2.7551E-06	1.6879E-04	2.9233E-05	1.5735E-05	4.1144E-05	60000
$f_{B,2}$	3.3063E-11	2.2156E-02	6.8156E-03	7.3960E-03	7.0154E-03	58997
$f_{B,3}$	1.2609E-09	3.9504E-06	1.9860E-07	1.4178E-08	7.3156E-07	60000
$f_{B,4}$	8.8130E-13	2.0416E-09	1.3088E-10	6.2361E-11	3.6564E-10	49251
$f_{B,5}$	1.5221E-12	1.5452E-09	1.3133E-10	6.3867E-11	2.8966E-10	50688
$f_{B,7}$	4.1898E+03	4.1898E+03	4.1898E+03	4.1898E+03	3.9534E-11	60000

Cuadro 5.29: Resultados de COPULEDA sin ajuste de varianza en 10 dimensiones

$F(x)$	Mejor	Peor	Media	Mediana	D.E.	Eval.
$f_{B,1}$	6.1848E-08	7.4085E-06	9.7234E-07	3.3514E-07	1.6614E-06	60000
$f_{B,2}$	3.5197E-11	4.6707E-02	1.2559E-02	7.3960E-03	1.2534E-02	54578
$f_{B,3}$	1.0260E-11	5.7731E-09	5.7284E-10	9.9731E-11	1.4079E-09	57088
$f_{B,4}$	6.0496E-13	2.5061E-10	7.3181E-11	7.2454E-11	5.3330E-11	46603
$f_{B,5}$	2.6404E-11	4.1257E-10	7.6998E-11	6.3649E-11	6.7443E-11	45053
$f_{B,7}$	6.2847E+03	6.2847E+03	6.2847E+03	6.2847E+03	3.8306E-10	60000

Cuadro 5.30: Resultados de COPULEDA sin ajuste de varianza en 15 dimensiones

$F(x)$	Mejor	Peor	Media	Mediana	D.E.	Eval.
$f_{B,1}$	1.0416E-06	4.5031E-05	6.3807E-06	4.1872E-06	8.0268E-06	60000
$f_{B,2}$	4.9568E-11	4.1797E-02	9.1931E-03	7.3960E-03	1.0454E-02	55875
$f_{B,3}$	2.1964E-10	2.0385E-07	1.7830E-08	4.5955E-09	3.8424E-08	60000
$f_{B,4}$	2.9787E-11	4.1706E-08	1.9996E-09	1.7101E-10	7.6267E-09	58222
$f_{B,5}$	4.4531E-11	6.7987E-09	8.3273E-10	2.1514E-10	1.5667E-09	57652
$f_{B,7}$	8.3796E+03	8.3796E+03	8.3796E+03	8.3796E+03	7.6555E-08	60000

Cuadro 5.31: Resultados de COPULEDA sin ajuste de varianza en 20 dimensiones

Dim.	Mejor	Peor	Media	Mediana	D.E.	Eval.
4	1.0103E+01	2.6369E+00	4.4186E+00	5.1025E+00	1.9704E+00	56915
30	5.0253E+00	2.5253E+00	2.9408E+00	2.5253E+00	9.4499E-01	60000

Cuadro 5.32: Resultados de COPULEDA sin ajuste de varianza para $f_{B,6}$.

Conjunto C

Los resultados del conjunto C en 5 dimensiones son mostrados en el cuadro 5.33, mientras que los resultados de la función $f_{C,1}$ en 10, 15 y 20 dimensiones

5.2. Resultados

se muestran en el cuadro 5.34.

$F(x)$	Mejor	Peor	Media	Mediana	D.E.	Eval.
$f_{C,1}$	9.9888E+04	3.2865E+02	3.3747E+04	9.9153E+03	3.7307E+04	60000
$f_{C,2}$	1.0105E+01	0.0000E+00	8.0841E+00	9.5999E+00	3.7057E+00	50183
$f_{C,3}$	1.0105E+01	0.0000E+00	6.9389E+00	9.0947E+00	3.4573E+00	55463

Cuadro 5.33: Resultados de COPULEDA sin ajuste de varianza en 5 dimensiones.

Dim.	Mejor	Peor	Media	Mediana	D.E.	Eval.
10	9.9158E+04	3.0568E+02	2.9892E+04	1.8492E+04	3.0866E+04	60000
15	9.4138E+04	8.1613E+02	2.5572E+04	1.8068E+04	2.4071E+04	60000
20	9.9451E+04	1.6589E+02	1.9720E+04	8.8041E+03	2.6884E+04	60000

Cuadro 5.34: Resultados de $f_{C,1}$ con COPULEDA sin ajuste de varianza.

5.2.4. COPULEDA con ajuste de varianza

Los resultados obtenidos por el COPULEDA con ajuste de varianza se muestran en los cuadros siguientes. El ajuste de varianza, al igual que en el caso del NOPREDA consiste en el procedimiento descrito en el capítulo anterior y además considera una mutación con un factor de 0.98 el cual disminuye de manera cuadrática conforme avanzan las generaciones. El número de individuos en la población es de 30 y el número máximo de evaluaciones de la función es de 60000.

Conjunto A

Los cuadros 5.35 a 5.38 muestran los resultados obtenidos por el COPULEDA con ajuste de varianza en 5,10, 15 y 20 dimensiones para el conjunto experimental A.

$F(x)$	Mejor	Peor	Media	Mediana	D.E.	Eval.
$f_{A,1}$	3.8605E-11	9.9483E-11	7.6142E-11	8.0466E-11	1.8574E-11	13328
$f_{A,2}$	2.7216E-11	9.9844E-11	7.1235E-11	7.3732E-11	2.2340E-11	23595
$f_{A,3}$	5.4533E-12	9.9271E-11	6.4679E-11	7.0665E-11	2.4930E-11	28322
$f_{A,4}$	2.8270E-11	9.8869E-11	7.7640E-11	8.0832E-11	2.0296E-11	18311
$f_{A,5}$	2.2706E-11	9.9779E-11	6.5785E-11	6.3695E-11	2.4691E-11	26585
$f_{A,6}$	1.9798E-11	9.8453E-11	7.6394E-11	8.2465E-11	2.1957E-11	23647
$f_{A,7}$	3.6030E-12	9.9845E-11	6.2976E-11	7.0683E-11	3.1335E-11	6771
$f_{A,8}$	6.1804E-04	1.5874E+00	2.6291E-01	1.8566E-02	4.8300E-01	60000
$f_{A,9}$	-1.0000E+10	-1.0000E+10	-1.0000E+10	-1.0000E+10	0.0000E+00	12493
$f_{A,10}$	-1.0000E+10	-1.0000E+10	-1.0000E+10	-1.0000E+10	2.6267E-06	57683

Cuadro 5.35: Resultados de COPULEDA con ajuste de varianza en 5 dimensiones.

5.2. Resultados

$F(x)$	Mejor	Peor	Media	Mediana	D.E.	Eval.
$f_{A,1}$	1.8721E-11	9.8281E-11	7.9491E-11	8.6603E-11	2.1556E-11	22261
$f_{A,2}$	2.0090E-11	9.9909E-11	7.8463E-11	8.7400E-11	2.2760E-11	33240
$f_{A,3}$	2.2812E-11	9.7979E-11	7.2782E-11	8.3473E-11	2.4949E-11	39779
$f_{A,4}$	6.9991E-12	9.9732E-11	8.0914E-11	8.8225E-11	2.2536E-11	30990
$f_{A,5}$	2.5685E-11	9.9879E-11	7.6081E-11	7.8982E-11	2.0660E-11	41335
$f_{A,6}$	1.7804E-11	9.9973E-11	7.2048E-11	7.5148E-11	2.2041E-11	34965
$f_{A,7}$	3.0669E-11	9.9196E-11	7.3181E-11	8.3628E-11	2.4026E-11	12298
$f_{A,8}$	1.0178E-04	1.6964E+00	4.5576E-01	1.6794E-01	5.3435E-01	60000
$f_{A,9}$	-1.0000E+10	-1.0000E+10	-1.0000E+10	-1.0000E+10	1.6230E-06	51611
$f_{A,10}$	-1.0000E+10	-1.0000E+10	-1.0000E+10	-1.0000E+10	5.6225E-06	60000

Cuadro 5.36: Resultados de COPULEDA con ajuste de varianza en 10 dimensiones.

$F(x)$	Mejor	Peor	Media	Mediana	D.E.	Eval.
$f_{A,1}$	5.5442E-11	9.9867E-11	8.6605E-11	8.9038E-11	1.1721E-11	31274
$f_{A,2}$	3.5428E-11	1.2868E-09	2.7697E-10	1.6814E-10	2.8441E-10	58932
$f_{A,3}$	7.8523E-09	4.5930E-07	8.4510E-08	6.1558E-08	8.7976E-08	60000
$f_{A,4}$	3.3174E-11	2.6113E-09	3.2521E-10	9.5164E-11	6.0776E-10	49935
$f_{A,5}$	6.7785E-10	8.7461E-07	6.3492E-08	9.7775E-09	1.7259E-07	60000
$f_{A,6}$	6.7512E-11	7.9028E-09	1.0620E-09	5.3665E-10	1.6889E-09	59814
$f_{A,7}$	3.5926E-11	9.9871E-11	8.0958E-11	8.1792E-11	1.5416E-11	16903
$f_{A,8}$	8.9736E-04	1.7355E+00	4.2037E-01	3.0367E-01	4.7521E-01	60000
$f_{A,9}$	-1.0000E+10	-1.0000E+10	-1.0000E+10	-1.0000E+10	1.9073E-06	60000
$f_{A,10}$	-1.0000E+10	-1.0000E+10	-1.0000E+10	-1.0000E+10	3.2910E-05	60000

Cuadro 5.37: Resultados de COPULEDA con ajuste de varianza en 15 dimensiones.

$F(x)$	Mejor	Peor	Media	Mediana	D.E.	Eval.
$f_{A,1}$	8.7419E-11	1.2548E-09	3.6326E-10	3.1452E-10	2.6882E-10	59875
$f_{A,2}$	8.5187E-08	4.2264E-06	8.8597E-07	5.2686E-07	1.0115E-06	60000
$f_{A,3}$	1.6310E-05	6.8730E-04	1.7654E-04	1.5566E-04	1.4350E-04	60000
$f_{A,4}$	2.4353E-10	4.4607E-06	3.6027E-07	6.6990E-09	9.3957E-07	60000
$f_{A,5}$	1.6865E-06	4.1438E-04	5.5421E-05	1.1596E-05	9.3063E-05	60000
$f_{A,6}$	1.0564E-07	1.1126E-06	4.5357E-07	3.7564E-07	2.8793E-07	60000
$f_{A,7}$	4.9399E-11	9.9934E-11	8.1969E-11	8.6126E-11	1.6588E-11	19138
$f_{A,8}$	1.0310E-03	3.8173E+00	8.4128E-01	4.2531E-01	9.4138E-01	60000
$f_{A,9}$	-1.0000E+10	-1.0000E+10	-1.0000E+10	-1.0000E+10	2.5293E-06	60000
$f_{A,10}$	-1.0000E+10	-1.0000E+10	-1.0000E+10	-1.0000E+10	1.8609E-03	60000

Cuadro 5.38: Resultados de COPULEDA con ajuste de varianza en 20 dimensiones.

Conjunto B

Al igual que en casos anteriores los resultados para la función $f_{B,6}$ se realizaron únicamente en 4 y 30 dimensiones y son mostrados en el cuadro 5.43. Los demás problemas del conjunto B se resolvieron en 5, 10, 15 y 20 dimensiones con el algoritmo del COPULEDA con ajuste de varianza y los resultados se

5.2. Resultados

muestran en los cuadros 5.39 a 5.42.

$F(x)$	Mejor	Peor	Media	Mediana	D.E.	Eval.
$f_{B,1}$	2.8248E-11	9.9899E-11	7.1217E-11	7.5092E-11	2.3036E-11	40283
$f_{B,2}$	1.1117E-11	3.4477E-02	1.4121E-02	1.3546E-02	9.5680E-03	53247
$f_{B,3}$	1.7166E-11	9.9305E-11	7.0127E-11	7.3807E-11	2.1717E-11	19698
$f_{B,4}$	2.4824E-11	9.6997E-11	7.5857E-11	8.3246E-11	1.8172E-11	15295
$f_{B,5}$	2.4085E-11	9.8911E-11	7.4063E-11	7.8023E-11	2.2311E-11	15651
$f_{B,7}$	2.0949E+03	2.0949E+03	2.0949E+03	2.0949E+03	3.1204E-04	4305

Cuadro 5.39: Resultados de COPULEDA con ajuste de varianza en 5 dimensiones

$F(x)$	Mejor	Peor	Media	Mediana	D.E.	Eval.
$f_{B,1}$	9.9221E-11	1.7236E-09	5.3423E-10	4.7366E-10	3.6692E-10	59897
$f_{B,2}$	7.8160E-11	5.6608E-02	1.6492E-02	1.3546E-02	1.3307E-02	53953
$f_{B,3}$	6.4943E-12	9.9973E-11	8.1748E-11	9.0771E-11	2.2851E-11	28440
$f_{B,4}$	3.9829E-11	9.9855E-11	8.5270E-11	9.3470E-11	1.7589E-11	22058
$f_{B,5}$	3.1708E-11	9.9941E-11	8.0830E-11	8.2823E-11	1.6296E-11	22350
$f_{B,7}$	4.1898E+03	4.1898E+03	4.1898E+03	4.1898E+03	4.5252E-05	7431

Cuadro 5.40: Resultados de COPULEDA con ajuste de varianza en 10 dimensiones

$F(x)$	Mejor	Peor	Media	Mediana	D.E.	Eval.
$f_{B,1}$	1.4354E-07	8.2208E-07	3.3371E-07	3.1028E-07	1.4614E-07	60000
$f_{B,2}$	5.2538E-11	2.2126E-02	5.6641E-03	1.4744E-09	6.8151E-03	43316
$f_{B,3}$	5.4825E-11	5.3407E-10	1.2708E-10	9.6505E-11	9.0946E-11	55077
$f_{B,4}$	3.2612E-11	9.9707E-11	8.7878E-11	9.2862E-11	1.4197E-11	35633
$f_{B,5}$	2.9098E-11	9.9902E-11	8.2443E-11	8.8519E-11	1.6550E-11	37296
$f_{B,7}$	6.2847E+03	6.2847E+03	6.2847E+03	6.2847E+03	1.3202E-04	10121

Cuadro 5.41: Resultados de COPULEDA con ajuste de varianza en 15 dimensiones

$F(x)$	Mejor	Peor	Media	Mediana	D.E.	Eval.
$f_{B,1}$	8.6008E-06	3.1151E-05	1.6886E-05	1.6037E-05	5.7004E-06	60000
$f_{B,2}$	8.8054E-11	1.9677E-02	5.9939E-03	7.3960E-03	6.2918E-03	57017
$f_{B,3}$	3.6836E-08	1.3553E-06	2.3863E-07	1.5758E-07	2.7524E-07	60000
$f_{B,4}$	1.3279E-09	2.3778E-08	5.2463E-09	4.4438E-09	4.6613E-09	60000
$f_{B,5}$	1.9590E-09	2.2878E-08	7.4933E-09	6.2228E-09	4.9154E-09	60000
$f_{B,7}$	8.3796E+03	8.3796E+03	8.3796E+03	8.3796E+03	2.1345E-05	16127

Cuadro 5.42: Resultados de COPULEDA con ajuste de varianza en 20 dimensiones

Dim.	Mejor	Peor	Media	Mediana	D.E.	Eval.
4	1.0103E+01	2.6369E+00	4.1533E+00	2.6880E+00	2.0283E+00	56157
30	5.0253E+00	2.5253E+00	2.6916E+00	2.5253E+00	6.3280E-01	60000

Cuadro 5.43: Resultados de COPULEDA con ajuste de varianza para $f_{B,6}$.

Conjunto C

En el cuadro 5.44 se muestran los resultados del COPULEDA con ajuste de varianza en 5 dimensiones para todos los problemas del conjunto experimental C, mientras que en el cuadro 5.45 se muestran los resultados del problema $f_{C,1}$ en 10, 15 y 20 dimensiones.

$F(x)$	Mejor	Peor	Media	Mediana	D.E.	Eval.
$f_{C,1}$	1.0000E+05	1.0000E+05	1.0000E+05	1.0000E+05	1.1639E-05	60000
$f_{C,2}$	1.0105E+01	9.0947E+00	9.6000E+00	9.6000E+00	5.1387E-01	35573
$f_{C,3}$	1.0105E+01	0.0000E+00	8.2863E+00	9.0947E+00	2.5647E+00	38360

Cuadro 5.44: Resultados de COPULEDA con ajuste de varianza en 5 dimensiones.

Dim.	Mejor	Peor	Media	Mediana	D.E.	Eval.
10	1.0000E+05	9.9705E+04	9.9990E+04	1.0000E+05	5.3837E+01	60000
15	1.0000E+05	9.9999E+04	1.0000E+05	1.0000E+05	1.2179E-02	60000
20	1.0000E+05	9.9522E+04	9.9983E+04	9.9999E+04	8.7064E+01	60000

Cuadro 5.45: Resultados de $f_{C,1}$ con COPULEDA y ajuste de varianza.

5.3. Comparación estadística del comportamiento de los algoritmos

La comparación entre el NOPREDA y el COPULEDA con y sin ajuste se realizó mediante pruebas de hipótesis sobre muestras de 30 resultados generados por los algoritmos. Como en este tipo de procesos se desconoce tanto la distribución como la población de la que provienen las muestras, el procedimiento comprendió la ejecución de una prueba *bootstrap* no paramétrica. El modelo general *bootstrap* se resume en el algoritmo 5.

Algoritmo 5 *Bootstrap*

- 1: Dada una muestra $\mathbf{x} = (x_1, x_2, \dots, x_k)$ calcular el estadístico $\hat{\theta}$.
 - 2: $b=0$
 - 3: **while** $b \leq B$ **do**
 - 4: Obtener una muestra con reemplazo de la muestra original para obtener $\mathbf{x}^{*b} = (x_1^{*b}, x_2^{*b}, \dots, x_k^{*b})$.
 - 5: De la muestra anterior calcular el estadístico $\hat{\theta}^{*b}$.
 - 6: $b=b+1$
 - 7: **end while**
 - 8: Emplee este estimador de la distribución de $\hat{\theta}$ para calcular la característica deseada.
-

Para este tipo de pruebas el valor de B debe ser grande (entre 999 a 9999). En el caso de la prueba *bootstrap* no paramétrica el remuestreo se toma a partir de la distribución empírica $\hat{F}_0(x)$ la cual satisface la hipótesis nula H_0 [31].

La hipótesis H_0 considera que ambas muestras son tomadas de la misma población (o bien que no existe una diferencia estadística relevante entre los algoritmos), mientras que la hipótesis alternativa H_A considera el caso contrario. Así pues, trabajando bajo H_0 , la muestra original del paso 1 estará compuesta por los 30 resultados generados por el primer algoritmo y por los 30 resultados del segundo. Es así que se realiza el procedimiento *bootstrap* con $x = (x_{1,1}, x_{1,2}, \dots, x_{1,30}, x_{2,1}, x_{2,2}, \dots, x_{2,30})$ como muestra original para obtener B muestras, de las que cuales se calculan los estadísticos $\hat{\theta}^{*b}$. Con estos estadísticos se puede esbozar la función de distribución y determinar si se rechaza o no la hipótesis nula con cierto nivel de significancia. Cuando la hipótesis alternativa considera que uno de los dos algoritmos es mejor que el otro, entonces se realiza una prueba de una cola. Supóngase que se tiene la siguiente prueba de hipótesis:

$$\begin{aligned} H_0 : \mu_1 &= \mu_2 \\ H_A : \mu_1 &< \mu_2. \end{aligned} \tag{5.1}$$

En este caso se considera que si bien se rechaza la hipótesis nula, entonces el primer algoritmo puede ser mejor que el segundo, luego entonces, el nivel de significancia será

$$P = p(T > t | H_0, \hat{F}_0) = \frac{1 + \#\{t^* \geq t\}}{B + 1}, \tag{5.2}$$

con $\#\{t^* \geq t\}$ como el número de veces que $t^* = \hat{\theta}^{*b}$ es mayor o igual a $t = \hat{\theta}$. Para esta prueba, el estadístico $\hat{\theta}$ es igual a la diferencia entre las medias de los resultados arrojados por cada algoritmo. Como en este caso $\mu_1 < \mu_2$, entonces $\hat{\theta} = \mu_2 - \mu_1$ y de forma similar para cada estadístico $\hat{\theta}^{*b}$.

Cuando no se tiene una hipótesis alternativa sobre cuál de los algoritmos es mejor, se puede realizar una prueba de dos colas, en la cual:

$$\begin{aligned} H_0 : \mu_1 &= \mu_2 \\ H_A : \mu_1 &\neq \mu_2, \end{aligned} \tag{5.3}$$

es la prueba a validar. Observe que ahora el nivel de significancia estará descrito por la expresión 5.4:

$$P = \frac{1 + \#\{t^* \geq t\} + \#\{t^* \leq -t\}}{B + 1}, \tag{5.4}$$

en cuyo caso es posible comparar el valor de P con la mínima significancia deseada para rechazar o no a H_0 .

A continuación se presentan los resultados obtenidos al emplear la prueba *bootstrap* no paramétrica en el NOPREDA y COPULEDA con y sin ajuste de varianza.

5.3.1. Resultados

Se realizaron comparaciones entre los algoritmos, primeramente para determinar si el ajuste de varianza generaba una mejora significativa en el algoritmo y luego, el mejor algoritmo del NOPREDA se comparó con el mejor del COPULEDA para así determinar el que mejor se desempeña en todos los conjuntos experimentales. Las pruebas se efectuaron con 30 corridas independientes de cada algoritmo, con y sin ajuste de varianza en todos los conjuntos experimentales y se empleó el valor absoluto de la diferencia entre las medianas como estadístico de prueba ($\hat{\theta} = |\tilde{\mu}_1 - \tilde{\mu}_2|$). El nivel de significancia se midió con la expresión 5.4 y el deseado se fijó en 5% con lo cual se aprobó o rechazó la hipótesis nula.

NOPREDA sin ajuste vs. NOPREDA con ajuste

En los cuadros 5.46, 5.58, 5.59 y 5.50 se muestran los resultados obtenidos luego de la prueba *bootstrap* no paramétrica con el NOPREDA sin ajuste de varianza y con ajuste de varianza en 5, 10, 15 y 20 dimensiones en casi todos los conjuntos experimentales. En el cuadro 5.47 además, se muestran los resultados de la comparación de estos mismos algoritmos para la función $f_{B,6}$ en 4 y 30 dimensiones. La hipótesis nula se rechazó si el nivel de significancia era menor del 5% en una prueba de dos colas y la mejor mediana determinó el mejor algoritmo.

$F(x)$	P-Valor	Mejor Mediana	Mejor Algoritmo
$f_{A,1}$	2.00E-03	8.21E-11	Con ajuste
$f_{A,2}$	1.50E-03	5.90E-11	Con ajuste
$f_{A,3}$	1.70E-03	6.40E-11	Con ajuste
$f_{A,4}$	1.50E-03	6.34E-11	Con ajuste
$f_{A,5}$	2.50E-03	4.23E-11	Con ajuste
$f_{A,6}$	2.10E-03	4.70E-11	Con ajuste
$f_{A,7}$	3.25E-01	2.59E-11	E.I.
$f_{A,8}$	2.00E-04	5.67E-01	Con ajuste
$f_{A,9}$	1.00E00	-1.00E+10	E.I.
$f_{A,10}$	1.00E00	-1.00E+10	E.I.
$f_{B,1}$	1.70E-03	8.30E-11	Con ajuste
$f_{B,2}$	3.97E-01	1.23E-02	E.I.
$f_{B,3}$	1.60E-03	7.10E-11	Con ajuste
$f_{B,4}$	1.90E-03	6.30E-11	Con ajuste
$f_{B,5}$	1.60E-03	7.39E-11	Con ajuste
$f_{B,7}$	1.00E-00	-2.09E+03	E.I.
$f_{C,1}$	3.00E-04	1.00E+05	Con ajuste
$f_{C,2}$	8.20E-03	1.01E+01	Con ajuste
$f_{C,3}$	2.06E-01	9.09E+00	Con ajuste

Cuadro 5.46: NOPREDA con ajuste vs. NOPREDA sin ajuste en 5 dimensiones.

E.I. significa ambos algoritmos se comportan Estadísticamente Igual o bien que en este caso no se rechaza la hipótesis nula $H_0 : \tilde{\mu}_1 = \tilde{\mu}_2$.

5.3. Comparación estadística del comportamiento de los algoritmos

Dim.	P-Valor	Mejor Mediana	Mejor Algoritmo
4	7.60E-01	5.10E+00	E.I.
30	2.90E-03	5.01E+00	Sin ajuste

Cuadro 5.47: NOPREDA con ajuste vs. NOPREDA sin ajuste en $f_{B,6}$.

$F(x)$	P-Valor	Mejor Mediana	Mejor Algoritmo
$f_{A,1}$	2.90E-03	9.05E-11	Con ajuste
$f_{A,2}$	3.40E-03	8.39E-11	Con ajuste
$f_{A,3}$	1.60E-03	8.36E-11	Con ajuste
$f_{A,4}$	2.83E-02	8.57E-11	Con ajuste
$f_{A,5}$	2.90E-03	6.06E-11	Con ajuste
$f_{A,6}$	2.80E-03	7.93E-11	Con ajuste
$f_{A,7}$	1.15E-01	4.99E-11	E.I.
$f_{A,8}$	2.30E-03	1.57E+00	Con ajuste
$f_{A,9}$	1.00E00	-1.00E+10	E.I.
$f_{A,10}$	1.00E00	-1.00E+10	E.I.
$f_{B,1}$	1.20E-03	1.44E-10	Con ajuste
$f_{B,2}$	2.52E-01	9.86E-03	E.I.
$f_{B,3}$	3.50E-03	9.21E-11	Con ajuste
$f_{B,4}$	2.55E-01	8.09E-11	E.I.
$f_{B,5}$	9.32E-02	8.85E-11	E.I.
$f_{B,7}$	1.00E00	-4.19E+03	E.I.
$f_{C,1}$	1.00E-04	9.95E+04	Con ajuste

Cuadro 5.48: NOPREDA con ajuste vs. NOPREDA sin ajuste en 10 dimensiones.

$F(x)$	P-Valor	Mejor Mediana	Mejor Algoritmo
$f_{A,1}$	1.09E-02	6.20E-11	Sin ajuste
$f_{A,2}$	3.20E-03	9.87E-11	Con ajuste
$f_{A,3}$	3.00E-03	1.53E-08	Con ajuste
$f_{A,4}$	1.22E-02	1.22E-10	Con ajuste
$f_{A,5}$	5.20E-03	9.89E-09	Con ajuste
$f_{A,6}$	2.60E-03	9.53E-11	Con ajuste
$f_{A,7}$	2.20E-03	3.40E-11	Sin ajuste
$f_{A,8}$	2.12E-02	1.66E+00	Con ajuste
$f_{A,9}$	1.00E00	-1.00E+10	E.I.
$f_{A,10}$	1.00E00	-1.00E+10	E.I.
$f_{B,1}$	9.00E-04	2.23E-07	Con ajuste
$f_{B,2}$	3.90E-03	9.74E-11	Con ajuste
$f_{B,3}$	2.70E-03	7.72E-11	Con ajuste
$f_{B,4}$	6.85E-01	9.19E-11	E.I.
$f_{B,5}$	3.80E-01	7.43E-11	E.I.
$f_{B,7}$	1.00E00	-6.28E+03	E.I.
$f_{C,1}$	1.00E-04	9.95E+04	Con ajuste

Cuadro 5.49: NOPREDA con ajuste vs. NOPREDA sin ajuste en 15 dimensiones.

5.3. Comparación estadística del comportamiento de los algoritmos

$F(x)$	P-Valor	Mejor Mediana	Mejor Algoritmo
$f_{A,1}$	3.59E-01	9.82E-11	E.I.
$f_{A,2}$	1.60E-03	2.62E-07	Con ajuste
$f_{A,3}$	1.38E-01	3.28E-05	E.I.
$f_{A,4}$	7.50E-03	1.18E-07	Con ajuste
$f_{A,5}$	4.40E-03	1.88E-05	Con ajuste
$f_{A,6}$	2.50E-03	1.78E-07	Con ajuste
$f_{A,7}$	4.18E-02	6.51E-11	Sin ajuste
$f_{A,8}$	2.13E-02	1.52E+00	Con ajuste
$f_{A,9}$	1.00E00	-1.00E+10	E.I.
$f_{A,10}$	1.00E00	-1.00E+10	E.I.
$f_{B,1}$	4.95E-01	9.06E-06	E.I.
$f_{B,2}$	6.51E-01	7.40E-03	E.I.
$f_{B,3}$	1.90E-01	3.60E-08	E.I.
$f_{B,4}$	2.00E-04	1.85E-10	Sin ajuste
$f_{B,5}$	1.00E-04	4.27E-10	Sin ajuste
$f_{B,7}$	1.00E00	-8.38E+03	E.I.
$f_{C,1}$	1.00E-04	1.00E+05	Con ajuste

Cuadro 5.50: NOPREDA con ajuste vs. NOPREDA sin ajuste en 20 dimensiones.

COPULEDA sin ajuste vs. COPULEDA con ajuste

En los cuadros 5.51, 5.53, 5.54 y 5.55 se muestran los resultados de la prueba *bootstrap* no paramétrica para el COPULEDA sin ajuste y con ajuste de varianza en 5, 10, 15 y 20 dimensiones. En el cuadro 5.52 se muestran asimismo, los resultados para la función $f_{B,6}$ en 4 y 30 dimensiones.

$F(x)$	P-Valor	Mejor Mediana	Mejor Algoritmo
$f_{A,1}$	4.70E-03	8.05E-11	Con ajuste
$f_{A,2}$	2.90E-03	7.37E-11	Con ajuste
$f_{A,3}$	2.30E-03	7.07E-11	Con ajuste
$f_{A,4}$	2.60E-03	8.08E-11	Con ajuste
$f_{A,5}$	2.80E-03	6.37E-11	Con ajuste
$f_{A,6}$	2.10E-03	8.25E-11	Con ajuste
$f_{A,7}$	7.76E-02	4.07E-11	E.I.
$f_{A,8}$	6.61E-02	1.86E-02	E.I.
$f_{A,9}$	1.00E00	-1.00E+10	E.I.
$f_{A,10}$	1.00E00	-1.00E+10	E.I.
$f_{B,1}$	1.30E-03	7.51E-11	Con ajuste
$f_{B,2}$	4.23E-01	1.35E-02	E.I.
$f_{B,3}$	1.60E-03	7.38E-11	Con ajuste
$f_{B,4}$	3.53E-01	7.34E-11	E.I.
$f_{B,5}$	2.38E-01	6.30E-11	E.I.
$f_{B,7}$	1.00E00	-2.09E+03	E.I.
$f_{C,1}$	1.60E-03	1.00E+05	Con ajuste
$f_{C,2}$	6.50E-01	9.60E+00	E.I.
$f_{C,3}$	2.97E-01	9.09E+00	E.I.

Cuadro 5.51: COPULEDA con ajuste vs. COPULEDA sin ajuste en 5 dimensiones.

5.3. Comparación estadística del comportamiento de los algoritmos

Dim.	P-Valor	Mejor Mediana	Mejor Algoritmo
4	1.59E-01	5.10E+00	E.I.
30	1.00E00	2.53E+00	E.I.

Cuadro 5.52: COPULEDA con ajuste vs. COPULEDA sin ajuste en $f_{B,6}$.

$F(x)$	P-Valor	Mejor Mediana	Mejor Algoritmo
$f_{A,1}$	3.17E-01	7.71E-11	E.I.
$f_{A,2}$	2.00E-03	8.74E-11	Con ajuste
$f_{A,3}$	1.90E-03	8.35E-11	Con ajuste
$f_{A,4}$	2.80E-03	8.82E-11	Con ajuste
$f_{A,5}$	2.70E-03	7.90E-11	Con ajuste
$f_{A,6}$	2.70E-03	7.52E-11	Con ajuste
$f_{A,7}$	7.00E-04	3.53E-11	Sin ajuste
$f_{A,8}$	4.40E-02	1.68E-01	Con ajuste
$f_{A,9}$	1.00E00	-1.00E+10	E.I.
$f_{A,10}$	1.00E00	-1.00E+10	E.I.
$f_{B,1}$	8.00E-04	4.74E-10	Con ajuste
$f_{B,2}$	5.74E-01	1.36E-02	E.I.
$f_{B,3}$	3.00E-03	9.08E-11	Con ajuste
$f_{B,4}$	8.10E-03	6.24E-11	Sin ajuste
$f_{B,5}$	3.50E-02	6.39E-11	Sin ajuste
$f_{B,7}$	1.00E00	-4.19E+03	E.I.
$f_{C,1}$	1.60E-03	1.00E+05	Con ajuste

Cuadro 5.53: COPULEDA con ajuste vs. COPULEDA sin ajuste en 10 dimensiones.

$F(x)$	P-Valor	Mejor Mediana	Mejor Algoritmo
$f_{A,1}$	1.10E-03	5.07E-11	Sin ajuste
$f_{A,2}$	1.80E-03	1.68E-10	Con ajuste
$f_{A,3}$	2.10E-03	6.16E-08	Con ajuste
$f_{A,4}$	2.20E-03	9.52E-11	Con ajuste
$f_{A,5}$	3.70E-03	9.78E-09	Con ajuste
$f_{A,6}$	3.20E-03	5.37E-10	Con ajuste
$f_{A,7}$	7.00E-03	5.53E-11	Sin ajuste
$f_{A,8}$	1.26E-01	3.04E-01	E.I.
$f_{A,9}$	1.00E00	-1.00E+10	E.I.
$f_{A,10}$	1.00E00	-1.00E+10	E.I.
$f_{B,1}$	6.86E-01	3.10E-07	E.I.
$f_{B,2}$	9.60E-02	1.47E-09	Con ajuste
$f_{B,3}$	4.56E-01	9.65E-11	E.I.
$f_{B,4}$	4.30E-03	7.25E-11	Sin ajuste
$f_{B,5}$	1.05E-02	6.37E-11	Sin ajuste
$f_{B,7}$	1.00E00	-6.28E+03	E.I.
$f_{C,1}$	1.20E-03	1.00E+05	Con ajuste

Cuadro 5.54: COPULEDA con ajuste vs. COPULEDA sin ajuste en 15 dimensiones.

5.3. Comparación estadística del comportamiento de los algoritmos

$F(x)$	P-Valor	Mejor Mediana	Mejor Algoritmo
$f_{A,1}$	4.40E-03	9.60E-11	Sin ajuste
$f_{A,2}$	2.20E-03	5.27E-07	Con ajuste
$f_{A,3}$	3.60E-03	2.86E-05	Sin ajuste
$f_{A,4}$	3.20E-03	6.70E-09	Con ajuste
$f_{A,5}$	3.00E-03	1.16E-05	Con ajuste
$f_{A,6}$	2.10E-03	3.76E-07	Con ajuste
$f_{A,7}$	6.99E-02	6.75E-11	E.I.
$f_{A,8}$	6.58E-01	4.25E-01	E.I.
$f_{A,9}$	1.00E00	-1.00E+10	E.I.
$f_{A,10}$	1.00E00	-1.00E+10	E.I.
$f_{B,1}$	1.00E-04	4.19E-06	Sin ajuste
$f_{B,2}$	3.06E-01	7.40E-03	E.I.
$f_{B,3}$	5.00E-04	4.60E-09	Sin ajuste
$f_{B,4}$	1.60E-03	1.71E-10	Sin ajuste
$f_{B,5}$	8.00E-04	2.15E-10	Sin ajuste
$f_{B,7}$	1.00E00	-8.38E+03	E.I.
$f_{C,1}$	1.20E-03	1.00E+05	Con ajuste

Cuadro 5.55: COPULEDA con ajuste vs. COPULEDA sin ajuste en 20 dimensiones.

NOPREDA vs. COPULEDA

En los cuadros 5.56, 5.58, 5.59 y 5.60 se muestran los resultados obtenidos por la prueba *bootstrap* no paramétrica con los mejores resultados del COPULEDA y el NOPREDA en 5, 10, 15 y 20 dimensiones. En el cuadro 5.57 se muestran asimismo, los resultados para la función $f_{B,6}$ en 4 y 30 dimensiones.

$F(x)$	P-Valor	Mejor Mediana	Mejor Algoritmo
$f_{A,1}$	7.55E-01	8.05E-11	E.I.
$f_{A,2}$	2.76E-01	5.90E-11	E.I.
$f_{A,3}$	4.76E-01	6.40E-11	E.I.
$f_{A,4}$	1.03E-01	6.34E-11	E.I.
$f_{A,5}$	1.73E-01	4.23E-11	E.I.
$f_{A,6}$	4.40E-03	4.70E-11	NOPREDA
$f_{A,7}$	3.79E-01	2.59E-11	E.I.
$f_{A,8}$	1.68E-02	1.86E-02	E.I.
$f_{A,9}$	1.00E00	-1.00E+10	E.I.
$f_{A,10}$	1.00E00	-1.00E+10	E.I.
$f_{B,1}$	3.70E-01	7.51E-11	E.I.
$f_{B,2}$	4.14E-01	1.23E-02	E.I.
$f_{B,3}$	7.46E-01	7.10E-11	E.I.
$f_{B,4}$	4.39E-01	6.30E-11	E.I.
$f_{B,5}$	1.98E-01	2.15E-10	E.I.
$f_{B,7}$	1.00E00	-2.09E+00	E.I.
$f_{C,1}$	2.80E-03	1.00E+05	COPULEDA
$f_{C,2}$	3.57E-01	1.01E+01	E.I.
$f_{C,3}$	3.35E-01	9.09E+00	E.I.

Cuadro 5.56: NOPREDA vs. COPULEDA en 5 dimensiones.

5.3. Comparación estadística del comportamiento de los algoritmos

Dim.	P-Valor	Mejor Mediana	Mejor Algoritmo
4	7.66E-01	5.10E+00	E.I.
30	1.00E-04	5.01E+00	NOPREDA

Cuadro 5.57: NOPREDA vs. COPULEDA en $f_{B,6}$.

$F(x)$	P-Valor	Mejor Mediana	Mejor Algoritmo
$f_{A,1}$	4.23E-01	8.66E-11	E.I.
$f_{A,2}$	6.85E-01	8.39E-11	E.I.
$f_{A,3}$	9.79E-01	8.35E-11	E.I.
$f_{A,4}$	7.78E-01	8.57E-11	E.I.
$f_{A,5}$	5.06E-02	6.06E-11	E.I.
$f_{A,6}$	6.52E-01	7.52E-11	E.I.
$f_{A,7}$	2.76E-01	3.53E-11	E.I.
$f_{A,8}$	3.00E-04	1.57E+00	NOPREDA
$f_{A,9}$	1.00E00	-1.00E+10	E.I.
$f_{A,10}$	1.00E00	-1.00E+10	E.I.
$f_{B,1}$	3.00E-03	1.44E-10	NOPREDA
$f_{B,2}$	1.45E-01	9.86E-03	E.I.
$f_{B,3}$	7.23E-01	9.08E-11	E.I.
$f_{B,4}$	8.58E-02	6.24E-11	E.I.
$f_{B,5}$	2.57E-02	6.39E-11	COPULEDA
$f_{B,7}$	1.00E00	-4.19E+03	E.I.
$f_{C,1}$	2.40E-03	1.00E+05	COPULEDA

Cuadro 5.58: NOPREDA vs. COPULEDA en 10 dimensiones.

$F(x)$	P-Valor	Mejor Mediana	Mejor Algoritmo
$f_{A,1}$	4.67E-01	5.07E-11	E.I.
$f_{A,2}$	9.34E-02	9.87E-11	E.I.
$f_{A,3}$	5.20E-03	1.53E-08	NOPREDA
$f_{A,4}$	1.82E-01	9.52E-11	E.I.
$f_{A,5}$	9.70E-01	9.78E-09	E.I.
$f_{A,6}$	2.90E-03	9.53E-11	NOPREDA
$f_{A,7}$	3.14E-01	3.40E-11	E.I.
$f_{A,8}$	2.00E-04	3.04E-01	COPULEDA
$f_{A,9}$	1.00E00	-1.00E+10	E.I.
$f_{A,10}$	1.00E00	-1.00E+10	E.I.
$f_{B,1}$	2.46E-02	2.23E-07	NOPREDA
$f_{B,2}$	4.76E-01	9.74E-11	E.I.
$f_{B,3}$	1.30E-03	7.72E-11	NOPREDA
$f_{B,4}$	2.25E-02	7.25E-11	COPULEDA
$f_{B,5}$	3.99E-01	6.37E-11	E.I.
$f_{B,7}$	1.00E00	-6.28E+03	E.I.
$f_{C,1}$	2.00E-04	1.00E+05	E.I.

Cuadro 5.59: NOPREDA vs. COPULEDA en 15 dimensiones.

5.4. Conclusiones del capítulo 5

$F(x)$	P-Valor	Mejor Mediana	Mejor Algoritmo
$f_{A,1}$	5.83E-01	9.60E-11	E.I.
$f_{A,2}$	5.80E-03	2.62E-07	NOPREDA
$f_{A,3}$	7.27E-01	2.86E-05	E.I.
$f_{A,4}$	1.73E-01	6.70E-09	E.I.
$f_{A,5}$	5.97E-01	1.16E-05	E.I.
$f_{A,6}$	1.56E-02	1.78E-07	E.I.
$f_{A,7}$	8.57E-01	6.51E-11	E.I.
$f_{A,8}$	5.20E-03	4.25E-01	COPULEDA
$f_{A,9}$	1.00E00	-1.00E+10	E.I.
$f_{A,10}$	1.00E00	-1.00E+10	E.I.
$f_{B,1}$	4.00E-03	4.19E-06	COPULEDA
$f_{B,2}$	7.04E-01	7.40E-03	E.I.
$f_{B,3}$	9.70E-03	4.60E-09	COPULEDA
$f_{B,4}$	8.78E-01	1.71E-10	E.I.
$f_{B,5}$	1.99E-01	2.15E-10	E.I.
$f_{B,7}$	1.00E00	-8.38E+03	E.I.
$f_{C,1}$	6.00E-03	1.00E+05	COPULEDA

Cuadro 5.60: NOPREDA vs. COPULEDA en 20 dimensiones.

En los resultados mostrados anteriormente se puede observar lo siguiente:

1. El desempeño de ambos algoritmos se mejora considerablemente luego de un ajuste de varianza en todas las dimensiones, aunque en el caso del NOPREDA se aprecia este efecto de forma más contundente que en el COPULEDA, ya que en este último, las diferencias entre los resultados del algoritmo con ajuste y sin ajuste, disminuyen conforme aumentan las dimensiones.
2. Existen funciones que se les dificultan tanto al NOPREDA como al COPULEDA, de tal suerte que para estos problemas, el desempeño de ambos es similar con y sin ajuste de varianza.
3. Al comparar el mejor algoritmo del NOPREDA con el mejor del COPULEDA se observa que ambos se desempeñan de forma equiparable a un nivel de significancia del 5%.

5.4. Conclusiones del capítulo 5

En el presente capítulo se mostraron los conjuntos experimentales con los cuales se probaron los algoritmos del NOPREDA y COPULEDA con y sin ajuste de varianza en varias dimensiones. Los problemas resueltos fueron tanto unimodales como multimodales y en casi todos los problemas se obtuvieron resultados satisfactorios. Posteriormente, se realizó una prueba *bootstrap* no paramétrica para determinar el desempeño de los algoritmos. Se compararon los resultados de esta prueba para cada algoritmo en cada una de sus versiones

5.4. Conclusiones del capítulo 5

y se observó que el ajuste de varianza mejora el comportamiento de ambos algoritmos y que en general tanto el NOPREDA como el COPULEDA, en su versión con ajuste de varianza, se desempeñan de forma similar.

5.4. Conclusiones del capítulo 5

Capítulo 6

Conclusiones generales

A continuación se presentan algunas de las conclusiones más importantes de este trabajo de investigación, así como una revisión general del contenido del mismo. Para observaciones más precisas, refiérase a cada una de las conclusiones que se encuentran al final de los capítulos.

Este trabajo de investigación se ha enfocado en el estudio de los diversos algoritmos de optimización evolutiva que se han propuesto a lo largo de los años, como es el caso de los algoritmos genéticos vistos en el capítulo uno, los cuales emplean para su desarrollo la idea evolutiva de la naturaleza conformada por selección, cruce y mutación. En ellos es posible ver que el óptimo es alcanzado mediante refinamientos de los mejores individuos de cada población durante cientos o miles de generaciones. Las técnicas empleadas para cada algoritmo varían en cada caso así como las heurísticas elegidas para la formación de cada operador genético, pero en todos se destaca el operador de cruce, como elemento principal de los algoritmos genéticos y las características de éstos que llevaron al estudio de una nueva generación de algoritmos evolutivos. Dichas características se resumen en una poca utilización de las dependencias entre las variables para solucionar problemas de índole deceptoria, lo cual atrajo la atención hacia modelos cuya principal característica es el modelado de la distribución de probabilidad de los buenos individuos para generar la nueva población. Este tipo de algoritmos, vistos en el capítulo dos, son conocidos como algoritmos de estimación de distribución y van desde los más sencillos, los cuales suponen independencia entre las variables hasta los más complejos, que toman en cuenta dependencias de orden 3 o mayor.

En los algoritmos de estimación de distribución, la etapa más importante es la forma como se modela la distribución de la selección para regenerar los nuevos individuos. En este tipo de algoritmos, a diferencia de los algoritmos genéticos, no se cuenta con operadores de cruce ni mutación, si no que se hace un estudio estadístico de los mejores individuos de la población actual para inducir las relaciones de las variables a las nuevas generaciones. Esta característica puede imprimirle una mala distribución a los individuos regenerados si es que se hace

una suposición previa sobre la distribución, sin conocer a ciencia cierta la forma como se relacionan las variables. Suponer una distribución, por otro lado, facilita enormemente las evaluaciones estadísticas y procura que el algoritmo evolucione de forma fluida. Esto podría resultar contradictorio, sin embargo, en los algoritmos propuestos hasta ahora, se coincide en las limitaciones para monitorizar las dependencias entre las variables y aproximarlas a una distribución previamente establecida.

En el capítulo 3 se presenta una forma de capturar las dependencias entre todas las variables e inducir las a un nuevo conjunto de individuos. Esta forma se incluye en un algoritmo de estimación de distribución conocido como el NOPREDA (Algoritmo de Estimación de Distribución No Paramétrico), el cual utiliza una técnica propuesta por Iman y Conover en 1982, para capturar las dependencias multivariadas entre los individuos seleccionados y formar una matriz de indexación capaz de guardar el orden de las asociaciones e inducir las al conjunto remuestreado para formar la nueva población. Este método es uno de los pioneros en el área de algoritmos de estimación de distribución continua y no paramétrica que además toma en cuenta las dependencias entre todas las variables. Más aún, en el capítulo 4 se introducen algunas de las definiciones y propiedades más importantes para explorar el concepto de *cópula*, el cual provee una forma alternativa de capturar las dependencias multivariadas al igual que lo hace el NOPREDA.

Dicho algoritmo es presentado como el COPULEDA (Algoritmo de Estimación de Distribución mediante Cópulas), el cual emplea la definición de *cópula* y el teorema de Sklar como punto medular en la medición de las dependencias entre las variables.

Pese a la gran diversidad de familias de *cópulas* existentes y estudiadas, el COPULEDA emplea la *cópula* gaussiana para medir las relaciones entre las variables, debido a sus propiedades y a lo simple de su implementación. Una de estas características, es que permite capturar las máximas dependencias (ya sean positivas o negativas) de las variables y puede extenderse al caso multivariado de una forma sencilla. Todo ello hace de la *cópula* gaussiana, la función ideal para su utilización en algoritmos de estimación de distribución. Los ejercicios realizados en el marco de este capítulo mostraron que:

1. Suponer independencia entre las variables para modelar la distribución de la que se remuestran los individuos, es en general insuficiente para realizar una buena aproximación a la distribución original.
2. Tanto el NOPREDA como el COPULEDA aproximan satisfactoriamente una distribución original dada, con lo cual los nuevos datos son generados con las relaciones deseadas entre las variables.

Además de presentar estos algoritmos, en el capítulo 4 se diseñó una forma de ajustar la varianza entre las variables para impedir la convergencia prematura a la que se confinan los algoritmos de estimación de distribución de este tipo, por la fiel representación de la distribución de la muestra a la que son sometidos.

Dicho ajuste se realiza de forma iterativa y no lineal tanto en el NOPREDA como en el COPULEDA.

Finalmente, en el capítulo 5 se realizaron los experimentos con un conjunto de varios problemas unimodales y multimodales en diversas dimensiones, para probar el desempeño de los algoritmos vistos tanto en su versión sin ajuste de varianza como en su versión con ajuste de varianza. Además, se incluyó una prueba *bootstrap* no paramétrica como una forma de realizar las comparaciones entre estos algoritmos y sus variantes de manera estadística con una confianza del 95 %. En todas estas pruebas se pudo observar que

1. El ajuste de la varianza mejoró el desempeño de ambos algoritmos en casi todas las pruebas y en casi todas las dimensiones, notándose una mejora significativa en el funcionamiento del NOPREDA con ajuste más que en el COPULEDA con ajuste.
2. Existen funciones que se les dificultan tanto al NOPREDA como al COPULEDA y el ajuste de varianza no muestra una mejora en el comportamiento de los algoritmos frente a dichas funciones. Éstas fueron diseñadas especialmente para algoritmos que emplean aglomeraciones en su búsqueda con lo cual el desempeño del NOPREDA y COPULEDA es pobre.
3. La comparación entre la mejor versión del NOPREDA contra la mejor del COPULEDA arrojó resultados muy interesantes que se pueden interpretar como igualdad en el funcionamiento de ambos algoritmos a un nivel de significancia del 5 %, y una mejora en los resultados de los problemas obtenidos por la mejor versión del COPULEDA en altas dimensiones.

Además de estas observaciones, no está por demás decir que, en efecto, las funciones de cópulas son elementos capaces de modelar las relaciones entre las variables y son aplicables a los algoritmos de estimación de distribución de forma sencilla y con resultados satisfactorios.

En este trabajo se mostró además, que la exploración de las técnicas estadísticas para su inclusión en los EDAs permanece como un área abierta para la investigación. Algo del trabajo a futuro que se mantiene en esta línea es:

1. Diseñar una heurística autoadaptable para el ajuste de las varianzas que no restrinja la exploración de los individuos y que impida la convergencia prematura hacia mínimos (o máximos) locales.
2. Estudiar el desempeño de las diversas familias de cópulas en los algoritmos de estimación de distribución, así como sus propiedades, las cuales pueden resultar beneficiosas para determinados problemas de optimización.
3. Formalizar nuevas técnicas que se puedan incluir en este tipo de algoritmos para resolver problemas que sólo funcionan con aquellos que incluyen técnicas de aglomeración en su procedimiento.

4. Desentrañar las razones teóricas y prácticas por las cuales a ciertas características de la población y de la medida de asociación se tienen problemas de singularidad que impiden hacer la factorización de Cholesky.

Referencias

- [1] Bäck, T. (1996) *Evolutionary Algorithms in Theory and Practice*. Oxford University Press.
- [2] Pelikan, M. (2005) *Hierarchical Bayesian Optimization Algorithm*. Springer.
- [3] Holland, J. (1975) *Adaptation in Natural and Artificial Systems*. University of Michigan Press.
- [4] Mitchell, M. (1996) *An introduction to Genetic Algorithms*. Massachusetts Institute of Technology Press.
- [5] Baluja, S. y Caruana, R. (1995) *Removing the Genetics from the Standard Genetic Algorithm*. Proceedings of the twelfth International Conference on Machine Learning.
- [6] Chow, C. y Liu, C. (1968) *Approximating Discrete Probability Distributions with Dependence Trees*. IEEE Transactions on Information Theory. Vol. IT-14, No. 3.
- [7] Soto, M. (2003) *Un Estudio sobre los Algoritmos Evolutivos con Estimación de Distribuciones basados en Poliárboles y su costo de evaluación*. Tesis Doctoral. Instituto de Cibernética, Matemática y Física de La Habana.
- [8] Larrañaga, P. y Lozano, J. (2002) *Estimation of Distribution Algorithms - A new Tool for Evolutionary Computation*. Kluwer Academic Publishers.
- [9] Tsutsui, S., Pelikan, M. y Goldberg, D. (2001) *Evolutionary Algorithm Using Marginal Histogram Models in Continuous Domain*. IlliGAL Report No. 2001019.
- [10] Sheskin, D. (1997) *Handbook of Parametric and Nonparametric Statistical Procedures*. CRC Press.
- [11] Joe, H. (1997) *Multivariate Models and Dependence Concepts*. Monographs on Statistics and Applied Probability 73. Chapman & Hall.

- [12] Iman, R. y Conover, J. (1982) *A Distribution-Free Approach to Inducing Rank Correlation Among Input Variables*. Communications in Statistics, Vol. B11, No. 3.
- [13] Iman, R. y Davenport, J. (1982) *Rank Correlation for use with Correlated Input Variables*. Communications in Statistics, Vol. B11, No. 3.
- [14] Jouanin, J., Riboulet, G. y Roncalli, T. *Financial Applications of Copula Functions*. Groupe de Recherche Opérationnelle, Crédit Lyonnais, France.
- [15] Schweizer, B. y Sklar, A. (1983). *Probabilistic Metric Spaces*. Elsevier, New York.
- [16] Yan, J. (2006). *Enjoy the Joy of Copulas*. Elsevier Science.
- [17] Trivedi P. y Zimmer, D. (2005). *Copula Modeling: An Introduction for Practitioners*, Foundations and Trends® in Econometrics, vol. 1, No. 1, pp 1-111.
- [18] Drouet, M. D. y Kotz, S. (2004) *Correlation and Dependence*. Imperial College Press.
- [19] Embrechts, P., McNeil, A. y Straumann, D. (2002). *Correlation and Dependence in Risk Management: Properties and Pitfalls*. In: M. A. H. Dempster(ed.): *Risk Management: Value at Risk and Beyond*. Cambridge: Cambridge University Press.
- [20] Nelsen, R. (1999). *An Introduction to Copulas*. Springer.
- [21] Shaw, W. y Lee K. (2007). *Copula Methods vs Canonical Multivariate Distributions: the multivariate Student T distribution with general degrees of freedom*. KCL Working Paper.
- [22] Frees, E. y Valdez E. (1997). *Understanding Relationships Using Copulas*. Originalmente presentado en el marco de la 32a. Conferencia de Investigación Actuarial del 6 al 8 de agosto de 1997 en la Universidad de Calgary. Calgary, Alberta, Canadá.
- [23] Durbin, J. y Stuart, A. (1951). *Inversions and Rank Correlations*. J. Roy. Statist. Soc. Ser. B 13.
- [24] Daniels, H. E. (1950). *Rank Correlations and Populations Models*. J. Roy. Statist. Soc. Ser. B 12.
- [25] Kruskal, W. H. (1958). *Ordinal Measures of Association*. J. Amer. Statist. Assoc. 53.
- [26] Genest, C. (1987). *Frank's Family of Bivariate Distributions*. Biometrika 74.
- [27] Abramowitz, M. y Stegun, I. (1972). *Handbook of Mathematical Functions*. New York: Dover, pp.927-933.

- [28] Grahl, J., Bosman P. y Rothlauf, F. (2006). *The Correlation-Triggered Adaptive Variance Scaling IDEA*. GECCO'06. Julio 8-12, 2006. Seattle, Washington, USA.
- [29] Zhang, Q., Sun, J., Tsang, E. y Ford, J. (2001). *Hybrid Estimation Algorithm for Global Optimization*. Department of Computer Science, University of Essex.
- [30] Lu, Q. y Yao, X. (2005). *Clustering and Learning Gaussian Distribution for Continuous Optimization*. IEEE Transactions on Systems, Man and Cybernetics-Part C: Applications and Reviews, Vol. 35, No. 2.
- [31] Davison, A. (1997). *Bootstrap Methods and their Applications*. Cambridge: Cambridge University Press. Cambridge Series in Statistical and Probabilistic Mathematics.

Referencias

Apéndice A

Forma paramétrica de ρ_S en la familia Clayton

Ya se ha visto que la medida de asociación ρ_S entre dos variable aleatorias X, Y se puede definir en términos de cópulas como

$$\rho_S = 12 \int_0^1 \int_0^1 C(u, v) dudv - 3, \quad (\text{A.1})$$

donde $C(u, v)$ es la función de cópula asociada al par X, Y . Establézcase entonces la función de cópula $C(u, v)$ de la familia Clayton como:

$$C(u, v) = (u^{-\alpha} + v^{-\alpha} - 1)^{-1/\alpha}, \text{ con } \alpha > 0. \quad (\text{A.2})$$

Al sustituir A.2 en la expresión A.1 se llega a una expresión cuya solución está dada por:

$$\rho_S = -3 - \frac{12 \text{MeijerG} \left\{ \left[\left(1 + \frac{1}{\alpha}, \frac{-2+\alpha}{\alpha}, \frac{-1+\alpha}{\alpha} \right), () \right], \left[\left(0, \frac{-1+\alpha}{\alpha} \right), \left(\frac{1}{\alpha} \right) \right], 1 \right\}}{\alpha \Gamma \left[\frac{1}{\alpha} \right]^2},$$

donde $\text{MeijerG} \{ [(a_1, \dots, a_n), (a_{n+1}, \dots, a_p)], [(b_1, \dots, b_m), (b_{m+1}, \dots, b_q)], z \}$ es una función definida como:

$$G_{qp}^{mn} \left(z \mid \begin{matrix} a_1, \dots, a_p \\ b_1, \dots, b_q \end{matrix} \right) = \frac{1}{2\pi i} \int z^{-s} \frac{\Gamma(1-a_1-s) \dots \Gamma(1-a_n-s) \Gamma(b_1+s) \dots \Gamma(b_m+s)}{\Gamma(a_{n+1}+s) \dots \Gamma(a_p+s) \Gamma(1-b_{m+1}-s) \dots \Gamma(1-b_q-s)} ds$$

y $\Gamma[z]$ es la función Gamma de Euler:

$$\Gamma[z] = \int_0^\infty t^{z-1} e^{-t} dt.$$