

Centro de Investigación en Matemáticas, A.C.

---

---

CIMAT

# EDAs con Funciones de Cópula

**T E S I S**

Que para obtener el grado de

**Maestro en Ciencias**

con Orientación en

**Ciencias de la Computación**

P r e s e n t a

**Emma Flores de la Fuente**

Director de Tesis:

**Dr. Arturo Hernández Aguirre**

Guanajuato, Gto. Septiembre de 2009

# Agradecimientos

Gracias...

A mis padres, Cristina y Rufino, por su amor, apoyo y confianza que siempre me han demostrado, sin ellos esto no sería posible.

A mis hermanos: Lucy, Vero, Hugo, Edith, Elisa y Ericka; cuñados y sobrinos, con los que he pasado grandes momentos y por todo el amor que me han brindado.

A mi asesor, el Dr. Arturo Hernández Aguirre, por su apoyo, paciencia, comprensión, consejos y sobre todo el tiempo que dedicó para poder sacar adelante este proyecto.

Agradezco también a los revisores, la Dra. Eloísa Días Francés, el Dr. Enrique Villa Diharce y el Dr. Salvador Ruiz Correa, por su tiempo dedicado y por las aportaciones que hicieron para mejorar este trabajo.

A David, Gustavo, Juan Carlos, Mario, Memo, Miguel y Valentín, por estos dos años que pasamos y por la ayuda que me brindaron.

A Judith y Ángel, por brindarme su amistad sincera.

Gracias a todos los que de alguna forma intervinieron en esta etapa de mi vida, recordaré los buenos momentos que facilitaron mi estancia en Guanajuato.

A CONACYT por el apoyo económico que me otorgó.

...Gracias a todos...Emma

# Índice general

<b>1. Introducción</b>	<b>1</b>
<b>2. Medidas de dependencia</b>	<b>6</b>
2.1. Correlación producto-momento . . . . .	7
2.1.1. Estimador de la correlación producto-momento. . . . .	8
2.2. Correlación por rangos . . . . .	9
2.3. $\rho_r$ de Spearman . . . . .	9
2.3.1. Estimador de $\rho_r$ de Spearman. . . . .	10
2.4. $\tau$ de Kendall . . . . .	11
2.4.1. Estimador de $\tau$ de Kendall. . . . .	12
2.5. Propiedades de las medidas de correlación . . . . .	12
2.6. Entropía e Información Mutua . . . . .	13
2.6.1. Entropía . . . . .	13
2.6.2. Entropía diferencial . . . . .	14
2.6.3. Entropía conjunta y condicional . . . . .	16
2.6.4. Entropía relativa . . . . .	16
2.6.5. Información mutua. . . . .	17
<b>3. EDAs con variables reales</b>	<b>19</b>
3.1. ¿Qué son los EDAs? . . . . .	20
3.2. Clasificación de los EDAs . . . . .	21
3.3. MIMIC . . . . .	25
<b>4. Funciones de Cópula</b>	<b>27</b>
4.1. ¿Qué son las Cópulas? . . . . .	27
4.2. Teorema de Sklar . . . . .	31
4.2.1. Función de densidad de una cópula . . . . .	31
4.2.2. Algunas Cópulas comunes. . . . .	32

4.2.3. Medidas de dependencia para cópulas . . . . .	38
<b>5. Inferencia Estadística para cópulas</b>	<b>41</b>
5.1. Método Exacto de Máxima Verosimilitud . . . . .	41
5.2. Método IFM . . . . .	44
5.3. Método CML . . . . .	45
5.4. Ejemplos de estimación de parámetro para cópulas . . . . .	46
5.5. Selección de cópula . . . . .	51
5.5.1. Cópula empírica . . . . .	51
<b>6. Simulación</b>	<b>56</b>
6.1. Muestreo condicional . . . . .	56
6.1.1. Ejemplos . . . . .	57
6.2. Muestreo elíptico . . . . .	63
<b>7. EDAs con funciones de cópula</b>	<b>67</b>
7.1. Planteamiento . . . . .	67
7.2. MIMIC con cópulas . . . . .	68
7.3. MIMIC con selección de cópula . . . . .	70
7.4. Experimentos y resultados . . . . .	74
7.4.1. Funciones de Prueba . . . . .	74
7.4.2. Implementación y resultados . . . . .	80
<b>8. Mezclas finitas de Cópulas Gaussianas</b>	<b>87</b>
8.1. Algoritmo EM . . . . .	89
8.2. Algoritmo EM en mezclas de Cópulas Gaussianas . . . . .	90
8.3. Resultados . . . . .	92
<b>9. Conclusiones</b>	<b>96</b>
<b>Bibliografía</b>	<b>98</b>

# Índice de cuadros

4.1. Medidas de dependencia para cópulas. . . . .	40
5.1. Función Log-Verosimilitud . . . . .	50
5.2. Función de Score. . . . .	51
6.1. Muestreo condicional para cópulas. . . . .	62
7.1. MIMIC-CG: Valor de Aptitud. . . . .	82
7.2. MIMIC-CG: Evaluaciones de función. . . . .	83
7.3. MIMIC-SC: Valor de Aptitud. . . . .	84
7.4. MIMIC-SC: Evaluaciones de función. . . . .	85
7.5. MIMIC-CG vs MIMIC-SC . . . . .	86
8.1. Mezcla de cópulas Gaussianas. . . . .	94

# Índice de figuras

1.1. Modelos Gráficos Probabilísticos. . . . .	3
2.1. Entropía . . . . .	14
2.2. Entropía diferencial . . . . .	15
2.3. Información mutua . . . . .	18
3.1. EDAs . . . . .	20
3.2. EDAs: Independientes. . . . .	23
3.3. EDAs: Bivariados . . . . .	24
3.4. EDAs: Dependencias múltiples . . . . .	25
4.1. Cota inferior de Fréchet-Hoeffding . . . . .	30
4.2. Cota superior de Fréchet-Hoeffding . . . . .	30
4.3. Cópula Producto . . . . .	35
4.4. Cópula de Ali-Mikhail-Haq . . . . .	36
4.5. Cópula Clayton . . . . .	36
4.6. Cópula de Farlie-Gumbel-Morgenstern . . . . .	37
4.7. Cópula Frank . . . . .	37
4.8. Cópula Gaussiana . . . . .	38
4.9. Cópula Gumbel . . . . .	38
6.1. Muestreo Cópula Ali-Mikhail-Haq. . . . .	64
6.2. Muestreo Cópula Clayton. . . . .	64
6.3. Muestreo Cópula Clayton. . . . .	65
6.4. Muestreo Cópula FGM. . . . .	65
6.5. Muestreo Cópula Frank. . . . .	66
6.6. Muestreo Cópula Gaussiana. . . . .	66
7.1. Funciones de prueba en dimensión dos . . . . .	80

8.1. Mezcla de Cópulas Gaussianas. . . . .	92
8.2. Mezcla de Cópulas Gaussianas. . . . .	93
8.3. Prueba 1. Mezcla de Cópulas Gaussianas. . . . .	94
8.4. Prueba 1. Mezcla de Cópulas Gaussianas . . . . .	95
8.5. Prueba 2. Mezcla de Cópulas Gaussianas . . . . .	95
8.6. Prueba 2. Mezcla de Cópulas Gaussianas . . . . .	95

# Lista de algoritmos

1.	Pseudocódigo para EDAs . . . . .	22
2.	Pseudocódigo para la estimación de la distribución de probabilidad conjunta para el MIMIC . . . . .	26
3.	Cópula empírica . . . . .	53
4.	Pseudocódigo para la selección de cópula . . . . .	54
5.	Muestreo condicional . . . . .	56
6.	Simulación de datos de una cópula Ali-Mikhail-Haq . . . . .	58
7.	Simulación de datos de una cópula Clayton . . . . .	59
8.	Simulación de datos de una cópula Farli-Gumbel-Morgenstern . . . . .	59
9.	Simulación de datos de una cópula Frank . . . . .	60
10.	Simulación de datos de una cópula Gumbel . . . . .	61
11.	Simulación de datos de una cópula Gaussiana . . . . .	63
12.	Pseudocódigo de un MIMIC con cópula Gaussiana . . . . .	71
13.	Pseudocódigo de un MIMIC con selección de cópula . . . . .	75
14.	Expectation Maximization . . . . .	89



# Capítulo 1

## Introducción

Los algoritmos evolutivos son métodos de optimización inspirados en los postulados de la evolución biológica. Están compuestos por un conjunto de entidades que representan posibles soluciones, las cuales se mezclan, y compiten entre sí, de tal manera que las más aptas son capaces de prevalecer a lo largo del tiempo, evolucionando hacia mejores soluciones. Los conceptos que utilizaremos al trabajar con los algoritmos evolutivos están relacionados con la terminología que se utilizan en el área de biología, el término *individuo o cromosoma* se refiere a las entidades que representan las soluciones al problema, además diremos que cada individuo está formado por genes, y al conjunto de individuos lo llamaremos *población*. Los individuos son modificados por operadores genéticos, principalmente selección, cruce y mutación. Los algoritmos evolutivos son utilizados principalmente en problemas con espacios de búsqueda extensos y no lineales, en donde otros métodos no son capaces de encontrar soluciones en un tiempo razonable.

Dentro de los algoritmos evolutivos se tiene el Algoritmo Genético (AG), que surgió en el año de 1970 de la mano de John Henry Holland [14]. Los algoritmos genéticos trabajan con una población, en la que cada individuo representa una solución factible a un problema dado. A cada individuo se le asigna un valor relacionado con la calidad de dicha solución, es decir, los más aptos tendrán una mejor puntuación, lo que les dará mayor probabilidad de ser seleccionados para la reproducción, y por otro lado los menos aptos tendrán una menor colaboración en la creación de la siguiente generación de individuos [24]. Un aspecto importante en los AG es la codificación, existen diferentes técnicas para llevar a cabo este proceso una de ellas consiste en representar a los individuos como cadenas binarias: secuencias de 1s y 0s,

donde cada entrada de la cadena es un gen.

Para que un AG tenga éxito, es necesario que en la población existan individuos que en su estructura contengan bloques del mejor individuo, para que al aplicar los operadores de cruce y mutación se tenga la posibilidad de construir la cadena óptima.

Un problema con el que los algoritmos genéticos tienen dificultades son los problemas con funciones de aptitud “engañosa” (deceptive function), en las que en cada iteración se intenta mejorar el valor de aptitud de los individuos pero a la vez se van perdiendo los bloques de la cadena óptima. Por ejemplo: imagine un problema en el que el espacio de búsqueda esté compuesto por todas las cadenas binarias de ocho caracteres cuyo óptimo es la cadena 11111111, y en el que la aptitud de cada individuo sea directamente proporcional al número de unos en él -es decir, 00000001 sería menos apto que 00000011, que sería menos apto que 00000111, etcétera -, con dos excepciones: la cadena 11111111 podría tener una aptitud muy baja, y la cadena 00000000 podría tener una aptitud muy alta. En este problema, un AG (al igual que la mayoría de los algoritmos) tendría probabilidad muy baja de encontrar un óptimo global.

Notemos que a los algoritmos genéticos no hacen uso de las relaciones que puedan existir entre los genes de un individuo, sino que solo dependen de los operadores genéticos. Aquí surge el problema de *linkage learning*, que consiste en identificar y preservar las interacciones importantes entre los genes. Muchas de las técnicas de *linkage learning* están orientadas a la identificación de subestructuras (bloques constructores) que puedan ser conservadas durante la mutación. Un algoritmo que aborda el problema del *linkage learning* es el Extended Compact Genetic Algorithm (ECGA) [13], que construye un modelo de probabilidad entre los genes para capturar sus interrelaciones y lo aplica al algoritmo genético convencional. Aunque se lograron mejoras al tratar problemas con funciones engañosas no fue suficiente para la captura de dependencia entre variables. Por ello, en los últimos años han surgido los Algoritmos de Estimación de Distribuciones (Estimation Distribution Algorithms, EDAs) como una alternativa con la que se ataca ese tipo de problemas, y que además tienen otras ventajas sobre los algoritmos genéticos, por ejemplo: no es necesaria la codificación de los individuos, disminuyen el número de parámetros ya que en éstos algoritmos la reproducción no se hace con base en los operadores de cruce y mutación, en lugar de ello, intentan capturar la dependencia entre las variables al estimar un **modelo** entre ellas. Las estrategias de modelado de los EDAs se basan en modelos gráficos

y probabilidades.

Los Modelos gráficos (ver [23]) son un enlace entre la teoría de probabilidad y la teoría de grafos, proporcionan un instrumento para tratar con dos problemas, la incertidumbre y la complejidad.

La idea de un modelo gráfico es interpretar un sistema complejo como la combinación de partes más simples, además proporcionan un modo simple de visualizar la estructura de un modelo probabilístico y pueden ser usados para diseñar y motivar nuevos modelos.

Un grafo está compuesto por nodos (o vértices) conectados por arcos (o aristas). En un modelo gráfico probabilístico, cada nodo representa una variable aleatoria y las aristas expresan relaciones probables entre las variables a las que está uniendo. Entre los modelos gráficos probabilísticos se tienen las redes Bayesianas, también conocidas como *Modelos Gráficos Dirigidos* en los que las aristas indican el sentido de dependencia entre las variables, el otro grupo es el de las redes de Markov o *Modelos Gráficos No Dirigidos*, en el que los arcos no indican dirección de dependencia entre variables. En la Figura 1.1 se presentan ejemplos de una modelos gráfico dirigido y uno no dirigido, en éste trabajo de tesis se eligen los modelos dirigidos y con base en tal modelo expresar su función de probabilidad subyacente.

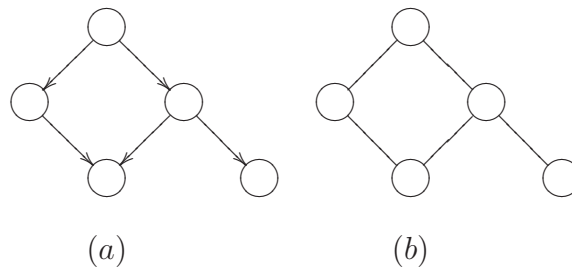


Figura 1.1: (a) Red Bayesiana (dirigida) (b) Red de Markov (no dirigida)

Los Modelos gráficos probabilísticos nos ayudarán a encontrar una estructura de dependencia entre las variables de nuestro problema para intentar capturar las interrelaciones entre ellas, pero eso no es suficiente ya que una vez capturada la dependencia nos gustaría inducirla en la siguiente generación.

Usualmente un algoritmo EDA utiliza un modelo gráfico para capturar la relación de dependencias y la estructura de los datos. Sin embargo, en esta tesis se modelan dichas dependencias por medio de funciones de cópula, que

son funciones con la propiedad de expresar la función de distribución conjunta de un grupo de variables en términos de sus distribuciones marginales además de que pueden conservar la correlación entre las variables.

En éste trabajo se hacen dos propuestas para trabajar con los EDAs y las Cópulas, la primera consiste en la construcción de un EDA tipo MIMIC (ver Capítulo 3) con cópula Gaussiana (EDA-CG) y la segunda es un EDA tipo MIMIC con Selección de Cópula (MIMIC-SG).

Dentro de los objetivos principales de éste trabajo de tesis se encuentran:

- . Aplicar los Algoritmos de Estimación de Distribuciones a problemas de optimización, intentando mejorar las soluciones que hasta ahora se han obtenido con algoritmos similares.
- . Presentar las funciones de cópula, explorar y explotar sus propiedades, ya que hemos considerado que pueden ser una buena alternativa para las funciones de distribución cuando se desea conservar cierto patrón de dependencia entre variables.
- . Proponer como modelo de un EDA una cópula Gaussiana, poder estimar su parámetro de dependencia y a partir de él generar una nueva población con el objetivo de mantener la correlación de la generación anterior y verificar si realmente se logran mejoras en su comportamiento.
- . Proponer un conjunto de cópulas y mediante un proceso de selección elegir la cópula que mejor capture las interrelaciones entre las variables.

El trabajo está organizado de la siguiente manera: en el Capítulo 2 se describen las medidas de dependencia y de concordancia para variables aleatorias, también se describen otras medidas como lo son la Entropía y la información mutua, que nos servirán para la implementación de el EDA con el que trabajamos. En el Capítulo 3 hablamos sobre los EDAs, se presentan sus características, su clasificación e introducimos el EDA tipo MIMIC con el que trabajaremos las funciones de cópula. En el Capítulo 4 se presentan las funciones de cópulas, aunque la teoría es muy extensa sólo presentamos los aspectos que consideramos de mayor relevancia. En el Capítulo 5 se presenta todo acerca de la inferencia estadística para cópulas, la estimación del parámetro y el método de selección de Copula propuesto por Ané y Kharoubi (2003) [2], que son las dos propuestas que hacemos para aplicar a los EDAs.

El Capítulo 6 está dedicado a los métodos de simulación de una cópula, en el Capítulo 7 se describen a detalle las dos propuestas que se hacen, y se presentan los resultados obtenidos con ambas implementaciones para cinco funciones de prueba. Por último, en el Capítulo 8 se plantea la idea de una mezcla finita de cópulas Gaussianas, inicialmente surgió como una opción para modelar muestras que se encuentran agrupadas en clusters, se aplica el algoritmo EM para estimar los parámetros de la mezcla y se presentan algunas pruebas que se hicieron.

Los programas fueron implementados en MATLAB *The Language of Technical Computing* Versión 7.0.0, la ventaja que se tiene es que cuenta con varias de las funciones estadísticas y numéricas que facilitan la implementación de las funciones de Cópula. Por otro lado, la desventaja que se tiene es el tiempo de ejecución, que en dimensión diez puede llegar a ser hasta de 30 horas para la función de Rosenbrock, que es una de las más complejas para optimizar.

# Capítulo 2

## Medidas de dependencia

Para comenzar con este trabajo de tesis presentamos algunas definiciones de conceptos sobre medidas de dependencia y de asociación que nos servirán como indicador del comportamiento de los valores de la muestra con los que estemos trabajando ya que nos interesa saber si existe algún tipo de relación entre las variables. Para el lector interesado en los temas presentados en éste capítulo puede consultar [11] y [20].

El concepto de independencia es fundamental en la Teoría de Probabilidad. Decimos que dos eventos  $A$  y  $B$  son independientes cuando la ocurrencia de uno de ellos no tiene influencia en el otro. En caso contrario son dependientes. De manera formal se tiene la siguiente definición,

**Definición 2.0.1.** *Las variables aleatorias  $X_1, \dots, X_n$  son independientes si para cualesquiera intervalos  $I_1, \dots, I_n$ , se cumple*

$$P[X_1 \in I_1, \dots, X_n \in I_n] = \prod_{i=1}^n P[X_i \in I_i]$$

Cuando se dice que las variables no son independientes, en realidad es mucho más complejo modelar su distribución conjunta. Como se mencionó anteriormente, lo que a nosotros nos interesa es saber qué tan relacionadas o dependientes son las variables con las trabajamos, entonces, comencemos presentando las medidas de correlación (lineal y por rangos).

## 2.1. Correlación producto-momento

El coeficiente de correlación producto-momento, también conocido como correlación de Pearson es un *estadístico paramétrico* que indica la fuerza y dirección de una relación lineal entre dos variables. De manera formal veamos la siguiente definición:

**Definición 2.1.1.** *La correlación producto-momento de variables aleatorias  $X, Y$  con esperanzas finitas  $E(X), E(Y)$  y varianzas finitas  $\sigma_X^2, \sigma_Y^2$  es,*

$$\rho(X, Y) = \frac{Cov(X, Y)}{\sigma_X \sigma_Y}.$$

donde  $Cov(X, Y)$  es la covarianza entre  $X$  y  $Y$  y se define como:

$$Cov(X, Y) = E[(X - \mu_X)(Y - \mu_Y)]$$

A la correlación producto-momento también se le conoce como *correlación lineal* o *correlación de Pearson*.

Es posible demostrar que la correlación lineal puede obtenerse mediante la siguiente expresión:

$$\rho(X, Y) = \frac{E(XY) - E(X)E(Y)}{\sigma_X \sigma_Y}.$$

La siguiente proposición es muy importante ya que establece que la medida de correlación lineal está acotada, lo que nos permitirá tener una referencia sobre que tan bien o están correlacionados los datos.

**Proposición 2.1.1.** *Para cualesquiera variables aleatorias  $X, Y$ , con medias y varianzas finitas se tiene que,*

$$-1 \leq \rho(X, Y) \leq 1.$$

Se puede afirmar que si  $X$  y  $Y$  son dos variables aleatorias independientes, entonces la correlación de Pearson es cero. Sin embargo, el recíproco en general no se cumple:

$$\begin{aligned} X \perp Y &\Rightarrow \rho(X, Y) = 0 \\ \rho(X, Y) = 0 &\not\Rightarrow X \perp Y \end{aligned}$$

El Lema 1 muestra que la correlación producto-momento se puede expresar en términos de la función de distribución conjunta de los vectores aleatorios  $(X, Y)$ .

**Lema 1.** *Sea  $(X, Y)$  un vector de variables aleatorias continuas con soporte  $\Omega = [a, b] \times [c, d]$  ( $a, b, c, d$  pueden ser infinitos). Sea  $F(x, y)$  la función de distribución conjunta acumulada de  $(X, Y)$  con distribuciones marginales acumuladas  $F_X(x)$  y  $F_Y(y)$ . Asumiendo que  $0 < \sigma_X^2, \sigma_Y^2 < \infty$  se tiene*

$$\rho(X, Y) = \frac{1}{\sigma_X \sigma_Y} \int \int_{\Omega} F(x, y) - F_X(x)F_Y(y) dx dy.$$

Es importante tener presente que la correlación no puede ser usada para inferir, en general, una relación causal entre variables, es decir, si se determina una correlación entre  $A$  y  $B$ , esto no significa que  $A$  causa a  $B$ . Podría existir otro conjunto de factores  $C$  que sea la causa de la relación entre  $A$  y  $B$ .

Es importante poder estimar la correlación producto-momento en base a una muestra ya que no siempre es posible saber la distribución teórica de la que proviene, por ello presentamos la forma de estimar el parámetro  $\rho$  en base a las observaciones.

### 2.1.1. Estimador de la correlación producto-momento.

Si se registran  $n$  observaciones de  $(X, Y)$ , el coeficiente de correlación producto-momento puede estimarse con cualquiera de las siguientes equivalencias:

$$\begin{aligned} \hat{\rho}(X, Y) &= \frac{\sum_{i=1}^n (x_i - \hat{\mu}_X)(y_i - \hat{\mu}_Y)}{(n-1)\hat{\sigma}_X\hat{\sigma}_Y} \\ &= \frac{n \sum x_i y_i - \sum x_i \sum y_i}{\sqrt{n \sum x_i^2 - (\sum x_i)^2} \sqrt{n \sum y_i^2 - (\sum y_i)^2}} \end{aligned}$$

Un inconveniente de la correlación de Pearson es que puede ser influenciada por outliers, no normalidad y no linealidad en los datos. La siguiente medida de concordancia está basada en el coeficiente de correlación de Pearson y en los rangos de los datos, pero que reduce considerablemente los posibles inconvenientes de no linealidad, no normalidad y la influencia de outliers en



la muestra.

## 2.2. Correlación por rangos

Antes de introducir el concepto de correlación por rangos es conveniente definir el concepto de concordancia.

Se dice que una pareja de datos  $(x_1, y_1)$  y  $(x_2, y_2)$  es concordante si

$$\operatorname{sgn}(x_2 - x_1) = \operatorname{sgn}(y_2 - y_1),$$

y es discordante en caso de que

$$\operatorname{sgn}(x_2 - x_1) = -\operatorname{sgn}(y_2 - y_1),$$

donde,

$$\operatorname{sgn}(x) = \begin{cases} 1 & \text{si } x > 0 \\ 0 & \text{si } x = 0 \\ -1 & \text{si } x < 0 \end{cases}$$

La correlación por rangos es una medida de concordancia basada en rangos, números de orden, de cada grupo de sujetos y compara dichos rangos. Spearman y Kendall propusieron, de manera separada, un coeficiente de correlación basado en los rangos de los valores de la muestra, veamos cada una de las propuestas.

## 2.3. $\rho_r$ de Spearman

La correlación de Spearman, o por rangos, fue introducida por Spearman (1904). Es una medida no paramétrica que consiste en reemplazar los valores originales de la muestra, de ambas variables, por números enteros positivos, comenzando del 1 en adelante, que correspondan a su ordenamiento de menor a mayor magnitud (rangos). Veamos una definición formal de la correlación por rangos.

**Definición 2.3.1.** *La correlación por rangos de las variables aleatorias  $X, Y$  con función de distribución acumulada  $F_X$  y  $F_Y$ , respectivamente, esta dada por*

$$\rho_r(X, Y) = \rho(F_X(X), F_Y(Y))$$

donde  $\rho$  es la correlación de Pearson.

Equivalentemente, en términos de concordancia se tiene la siguiente definición,

**Definición 2.3.2.** Si  $(X_1, Y_1)$  y  $(X_2, Y_2)$  son dos parejas de variables aleatorias independientes con función de distribución conjunta  $F_{XY}$  y distribuciones marginales  $F_X$  y  $F_Y$ , el coeficiente de correlación por rangos  $\rho_r$  de Spearman está definido de la siguiente manera

$$\rho_r = 3P[(X_1 - X_2)(Y_1 - Y_2) > 0] - P[(X_1 - X_2)(Y_1 - Y_2) < 0]. \quad (2.1)$$

Dado que trabajamos con muestras, es importante tener una expresión para estimar el parámetro  $\rho$  de Spearman.

### 2.3.1. Estimador de $\rho_r$ de Spearman.

Dada una muestra de variables aleatorias  $(x_i, y_i)$  de tamaño  $N$ , su correspondiente coeficiente de correlación por rangos se define como:

$$\rho_r(X, Y) = \frac{\sum_{i=1}^N (R_i - \bar{R})(S_i - \bar{S})}{\sqrt{\sum_{i=1}^N (R_i - \bar{R})^2 \sum_{i=1}^N (S_i - \bar{S})^2}}$$

donde,

$R_i$  es el rango de  $x_i$

$S_i$  es el rango de  $y_i$

$$\bar{R} = \frac{1}{N} \sum_{i=1}^N R_i$$

$$\bar{S} = \frac{1}{N} \sum_{i=1}^N S_i$$

Considerando que los rangos toman como valor los primeros  $n$  números enteros, el coeficiente de correlación por rangos puede expresarse como:

$$\rho_r(X, Y) = 12 \frac{\sum_{i=1}^N (R_i - \bar{R})(S_i - \bar{S})}{N(N^2 - 1)}$$

o

$$\rho_r(X, Y) = 1 - 6 \frac{\sum_{i=1}^N (R_i - S_i)^2}{N(N^2 - 1)}$$

El coeficiente de correlación de Spearman es recomendable utilizarlo cuando los datos presentan valores extremos ya que dichos valores afectan mucho el coeficiente de correlación de Pearson, o ante distribuciones no normales. La segunda medida de correlación por rangos es la propuesta por Kendall al igual que la propuesta por Spearman se basa en los rangos de la muestra, pero veamos su definición formal.

## 2.4. $\tau$ de Kendall

Es un estadístico *no paramétrico* usado para medir el grado de correspondencia entre dos rangos.

**Definición 2.4.1.** Si  $(X_1, Y_1)$  y  $(X_2, Y_2)$  son dos parejas de variables aleatorias independientes con función de distribución conjunta  $F_{XY}$  y distribuciones marginales  $F_X$  y  $F_Y$ , el coeficiente de correlación por rangos  $\tau$  de Kendall está definido de la siguiente manera

$$\begin{aligned}\tau &= P[(X_1 - X_2)(Y_1 - Y_2) > 0] - P[(X_1 - X_2)(Y_1 - Y_2) < 0]. \\ &= 2P[(X_1 - X_2) \geq 0] - 1\end{aligned}$$

Así,

$$\begin{aligned}P[(X_1 - X_2)(Y_1 - Y_2) > 0] &= P[X_1 > X_2, Y_1 > Y_2] \\ &\quad + P[X_1 < X_2, Y_1 < Y_2] \\ &= 2P[X_1 > X_2, Y_1 > Y_2].\end{aligned}$$

Si fijamos las variables  $(X_1, Y_1)$  en  $(x_1, y_1)$  se tiene que

$$P(X_2 < x_1, Y_2 < y_1) = F_{XY}(x_1, y_1)$$

y si ahora hacemos variar  $(X_1, Y_1)$  sobre su soporte, tenemos:

$$P(X_2 < X_1, Y_2 < Y_1) = \int_{\mathcal{X}} \int_{\mathcal{Y}} F_{XY} dF_{XY}$$

y regresando a la definición de  $\tau$  se llega a que la correlación por rangos de Kendall se puede calcular como,

$$\tau = 4 \int_{\mathcal{X}} \int_{\mathcal{Y}} F_{XY} dF_{XY} - 1. \quad (2.2)$$

La expresión 2.2 se utilizará posteriormente para determinar la correlación por rangos de una cópula, por ahora veamos como se puede estimar el parámetro  $\tau$  de Kendall para una muestra.

### 2.4.1. Estimador de $\tau$ de Kendall.

Si se considera una muestra de tamaño  $n$ , el coeficiente de correlación  $\tau$  de Kendall puede estimarse mediante la siguiente expresión,

$$\tau = \frac{n_c - n_d}{\frac{1}{2}n(n-1)}$$

donde,

$n_c$  es la cantidad de parejas concordantes y

$n_d$  es la cantidad de parejas discordantes.

En caso de haber empates, el cálculo de  $\tau$  de Kendall debe hacerse con el siguiente ajuste:

$$\tau = \frac{n_c - n_d}{\sqrt{n_c + n_d + n_y} \sqrt{n_c + n_d + n_x}}$$

donde

$n_c$  es la cantidad de parejas concordantes.

$n_d$  es la cantidad de parejas discordantes.

$n_y$  es la cantidad de empates en  $X$ , llamada también pareja *extra-y*.

$n_x$  es la cantidad de empates en  $Y$ , llamada también pareja *extra-x*.

El parámetro  $\tau$  de Kendall coincide con  $\rho$  de Spearman en ser una medida libre de distribución, pero eso no implica que sean idénticas en magnitud.

Hasta ahora solo hemos introducido los conceptos de correlación para cualesquiera funciones de distribución conjunta, pero en el Capítulo 4 se darán los coeficientes de correlación  $\rho_r$  y  $\tau$  en términos de cópulas.

## 2.5. Propiedades de las medidas de correlación

En [6] se presentan algunas propiedades que comparten las medidas de asociación que hemos descrito,  $\rho$  de Pearson,  $\rho_r$  de Spearman y  $\tau$  de Kendall, las cuales se enuncian a continuación. Para  $M_{X,Y}$  cualesquiera de las medidas de asociación previamente mencionadas se cumple que:

1.  $M_{X,Y} \in [-1, 1]$ .
2. Simetría:  $M_{X,Y} = M_{Y,X}$ .
3. Si  $X$  y  $Y$  son independientes, entonces  $M_{X,Y} = 0$ .
4.  $M_{-X,Y} = M_{X,-Y} = -M_{X,Y}$

Hemos visto los conceptos de correlación lineal y correlación por rangos, pero existen otras medidas que nos servirán como indicador entre dos variables aleatorias,  $X$  e  $Y$ , para saber qué tanta información o incertidumbre se puede conocer de la primera variable ( $X$ ), en términos de la otra variable ( $Y$ ).

## 2.6. Entropía e Información Mutua

### 2.6.1. Entropía

LA entropía de una variable aleatoria  $X$ , denotada por  $H(X)$ , fue introducida por Claude E. Shannon [31]. El concepto de entropía tiene que ver con la incertidumbre que existe en cualquier experimento o señal aleatoria. La entropía está relacionada con la cantidad de ruido o desorden que contiene o libera un sistema.

Si consideramos una variable aleatoria discreta  $X$  con distribución de probabilidad dada por  $P[X = x] = p(x)$ , entonces el cálculo de la entropía de  $X$  se realiza mediante la siguiente expresión:

$$\begin{aligned} H(x) &= \sum_{x \in \mathcal{X}} p(x) \log_b \left( \frac{1}{p(x)} \right) \\ &= - \sum_{x \in \mathcal{X}} p(x) \log_b(p(x)) \\ &= -E_{p(x)}[\log_b(p(x))] \end{aligned}$$

donde,  $\mathcal{X}$  es el soporte de la variable aleatoria  $X$ .

En la Figura 2.1 se muestra la entropía, tomando diferentes bases del logaritmo, de una variable aleatoria discreta  $X$  que sigue una distribución Bernoulli con parámetro  $p$ .

Pero si la variable aleatoria es continua, la definición de entropía que se ha dado no será válida, por lo que es necesario definir la entropía diferencial, que involucra variables aleatorias continuas.

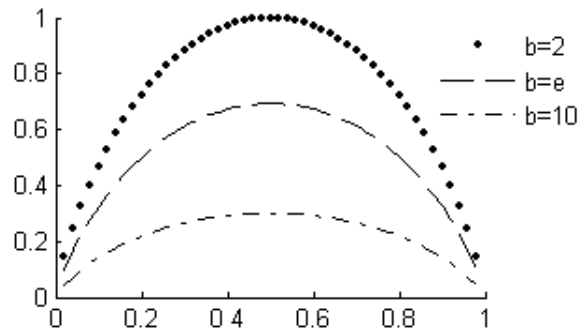


Figura 2.1:  $H(X) = p \log_b(p) + (1 - p) \log_b(1 - p)$ . Entropía de una variable aleatoria discreta utilizando tres diferentes bases de logaritmo.

### 2.6.2. Entropía diferencial

Si la variable aleatoria  $X$  es continua, la entropía recibe el nombre de **entropía diferencial**, y esta dada por la siguiente expresión,

$$\begin{aligned} H(x) &= \int_{\mathcal{X}} f(x) \log_b \left( \frac{1}{f(x)} \right) \\ &= - \int_{\mathcal{X}} f(x) \log_b(f(x)) \\ &= -E_{f(x)}[\log_b(f(x))] \end{aligned}$$

al igual que en el caso discreto,  $\mathcal{X}$  es el soporte de la variable aleatoria  $X$ , y  $f(x)$  representa la función de densidad de la variable aleatoria  $X$ . Para ejemplificar la entropía diferencial consideremos una variable aleatoria  $X \sim N(\mu, \sigma)$  cuya función de densidad está dada por:

$$p(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp \left\{ -\frac{1}{2} \left( \frac{x - \mu}{\sigma} \right)^2 \right\},$$

entonces su entropía diferencial es,

$$H(x) = - \int_{-\infty}^{\infty} \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right\} \log_b\left(\frac{1}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right\}\right) dx$$

haciendo los cálculos correspondientes se llega a que la entropía diferencial para una variable aleatoria que sigue una distribución normal con media  $\mu$  y varianza  $\sigma^2$  esta dada por:

$$H(x) = \frac{1}{2} \log_b(2\pi e\sigma^2)$$

En la figura 2.2 se muestra la entropía diferencial para una variable normal con diferentes bases para el logaritmo.

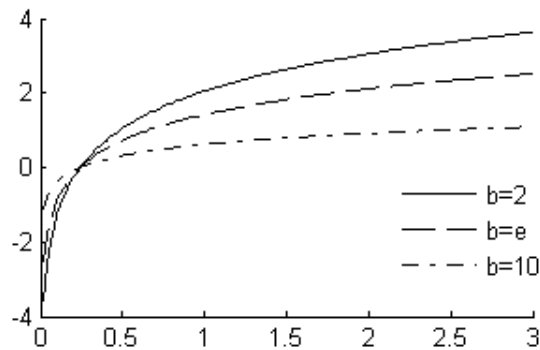


Figura 2.2: Entropía de una variable aleatoria continua que sigue una distribución normal utilizando tres diferentes bases de logaritmo.

Dado que se trabaja con distribuciones bivariadas, es decir, trabajamos las variables por parejas, es conveniente introducir el concepto de entropía conjunta, entropía relativa y entropía condicional.

### 2.6.3. Entropía conjunta y condicional

Comenzamos con la definición de la entropía conjunta. Considere 2 variables aleatorias  $X$  y  $Y$  con función de densidad conjunta  $p(x, y)$ , entonces su entropía conjunta está dada por,

$$H(X, Y) = -E_{p(x,y)}[\log_b(p(x, y))], \quad (2.3)$$

y su Entropía condicional queda determinada como,

$$H(X|Y) = -E_{p(x,y)}[\log_b(p(x|y))]. \quad (2.4)$$

Veamos con un ejemplo la entropía conjunta y la entropía relativa, cuando se tiene una distribución normal bivariada, ya que serán necesarias en éste trabajo en capítulos posteriores.

Ejemplo. Sean  $X, Y$  variables aleatorias con distribución normal bivariada, su entropía conjunta es,

$$H(X, Y) = \frac{1}{2} \log_b((2\pi e)^2 |\Sigma|)$$

donde,  $|\Sigma| = \sigma_X^2 \sigma_Y^2 - \sigma_{XY}^2$   
y su entropía condicional es,

$$H(X|Y) = \frac{1}{2} \log_b((2\pi e \sigma_{cond}^2))$$

donde,  $\sigma_{cond}^2 = \sigma_X^2 - \frac{\sigma_{XY}^2}{\sigma_Y^2}$

### 2.6.4. Entropía relativa

También conocida como divergencia de Kullback-Leibler, divergencia de la información o ganancia de la información, es una medida de discrepancia entre dos distribuciones de probabilidad; Por ejemplo, entre una distribución de probabilidad “verdadera”  $P$  a otra distribución de probabilidad arbitraria  $Q$  Ahora presentamos una definición para la divergencia Kullback-Leibler [32], pero si se desea se puede consultar [23] ó [19].

**Definición 2.6.1. Divergencia de Kullback-Leibler.** Para dos funciones de probabilidad  $P$  y  $Q$  de dos variables aleatorias discretas, la divergencia de Kullback-Leibler ( $K-L$ ) entre  $Q$  y  $P$  está definida como:

$$D_{KL}(P, Q) = \sum_{x \in \mathfrak{X}} P(x_i) \log_b \left( \frac{P(x_i)}{Q(x_i)} \right).$$



Para distribuciones  $P$  y  $Q$  de dos variables aleatorias continuas, las sumas se transforman en integrales de la siguiente manera:

$$D_{KL}(P, Q) = \int_{-\infty}^{\infty} p(x) \log_b \left( \frac{p(x)}{q(x)} dx \right).$$

Algunas de las propiedades de la divergencia K-L, son las siguientes:

1. Satisface la desigualdad de Gibb

$$D_{KL}(P, Q) \geq 0$$

2. En general, la divergencia K-L no es simétrica, es decir,

$$D_{KL}(P, Q) \neq D_{KL}(Q, P)$$

### 2.6.5. Información mutua.

Es una medida de dependencia entre dos variables aleatorias  $X$  y  $Y$ . Intuitivamente la información mutua mide cuanto conocimiento de una de las variables reduce la incertidumbre acerca de la otra. Para el caso discreto la información mutua esta dada por,

$$I(X; Y) = \sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{Y}} p(x, y) \log \left( \frac{p(x, y)}{p(x)p(y)} \right)$$

y para el caso continuo se tiene,

$$I(X; Y) = \int_{\mathcal{X}} \int_{\mathcal{Y}} p(x, y) \log \left( \frac{p(x, y)}{p(x)p(y)} \right)$$

donde  $\mathcal{X}$  y  $\mathcal{Y}$  son el soporte de las variables  $X$  y  $Y$  respectivamente.

Ejemplo. Dadas dos variables aleatorias con distribución normal bivariada, la información mutua es

$$I(X, Y) = -\frac{1}{2} \log_b(1 - \rho^2)$$

donde,  $\rho$  es la medida de correlación por rangos de Spearman, en la Figura 2.3 se muestra la información mutua para el caso de variables con distribución normal bivariada, utilizando diferentes bases para el logaritmo.

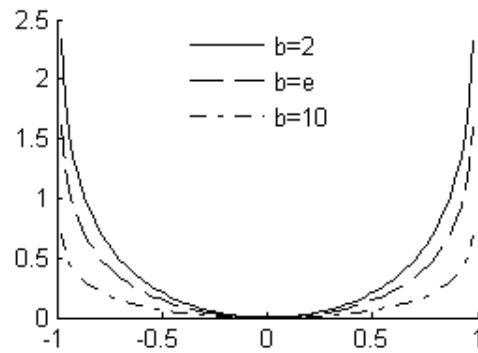


Figura 2.3: Información mutua para variables aleatorias continuas con distribución normal bivariada para tres diferentes bases de logaritmo.

Se han presentado las medidas que consideramos más importante para medir dependencia y concordancia entre variables, y son las que utilizaremos para la implementación del EDA con el que trabajaremos, pero aún no se ha hecho la representación oficial de lo que son los EDAs, y eso es lo que se hará en el siguiente capítulo.

# Capítulo 3

## EDAs con variables reales

El comportamiento de los algoritmos genéticos depende de un gran número de parámetros como el tamaño de la población, las probabilidades de cruce y mutación, etc. De hecho la determinación de valores adecuados para dichos parámetros constituye por si mismo un problema de optimización. Una mala elección de los valores de los parámetros puede llevar a que el algoritmo obtenga soluciones alejadas del valor óptimo. Este es uno de los motivos por los que desde hace varios años se han venido estudiando alternativas a los métodos heurísticos estocásticos existentes que no necesiten un número grande de parámetros por ajustar.

Otro motivo por el que se ha desarrollado la búsqueda de nuevos heurísticos de optimización es la necesidad de identificar las interrelaciones entre las variables utilizadas para representar a los individuos.

Todo ello ha motivado el nacimiento de un nuevo método de búsqueda conocido como Algoritmo de Estimación de Distribuciones, al que denotaremos como EDAs (*Estimation of Distribution Algorithms*).

Es necesario definir algunos términos que utilizaremos en este trabajo, el concepto *individuo* se refiere a cada una de las soluciones, generalmente es un vector de tamaño  $d \times 1$ , donde  $d$  representa la dimensión en la que se trabaja, su representación puede ser discreta o real, *población* es el conjunto de individuos, y *generación* es cada una de las poblaciones que evolucionan sucesivamente. Ahora veamos en que consisten los EDAs.

### 3.1. ¿Qué son los EDAs?

Los EDAs, fueron introducidos originalmente por Mühlenbein y Paab [26], están basados en poblaciones que evolucionan, pero tal evolución no se lleva a cabo por medio de los operadores de cruce y mutación como en los algoritmos genéticos, en lugar de ello la nueva población de individuos se muestrea de una distribución de probabilidad, la cual es estimada del conjunto de individuos seleccionados de entre los que constituyen la generación anterior, en la Figura 3.1 se presenta la idea en la que se basan los EDAs.

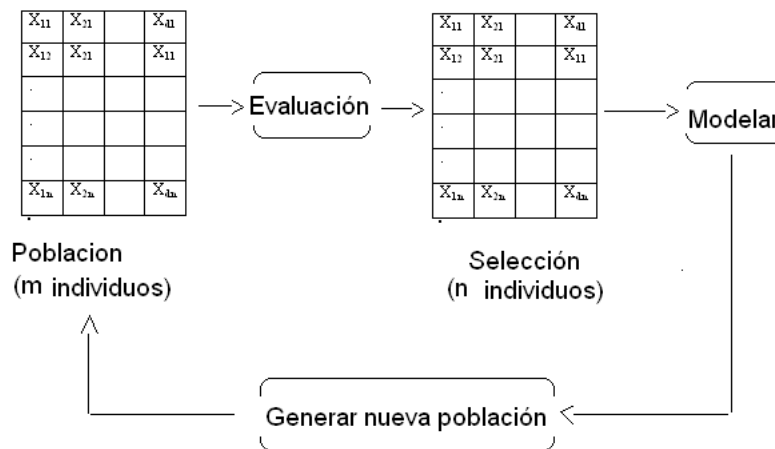


Figura 3.1: Ilustración de un EDA

La estimación de la distribución de probabilidad conjunta asociada a los individuos seleccionados en cada generación es la principal dificultad de esta aproximación.

Esta nueva clase de algoritmos también es conocida en la literatura como Probabilistic Model Building Genetic Algorithms, PMBGA [29] e Iterated Density Estimators Evolutionary Algorithms, IDEAS [5].

En el algoritmo 1 se presenta el pseudocódigo de un EDA en su forma básica, que sigue esencialmente los siguientes pasos:

1. Como primer paso, se genera una población inicial  $D_0$  formada por  $m$  individuos. Generalmente se asume que los individuos de la población

inicial siguen una distribución uniforme en cada variable. Posteriormente evaluamos cada individuo de  $D_0$  en la función objetivo.

2. Para  $k \geq 1$ , seleccionamos los  $n$  ( $n \leq m$ ) mejores individuos de  $D_{k-1}$  para formar un nuevo conjunto  $D_{k-1}^*$ .
3. Se induce un modelo probabilístico  $M$  que represente las interdependencias entre las  $d$  variables. Podemos decir que este paso es el más importante en el EDA, ya que la siguiente generación y la evolución satisfactoria hacia mejores individuos depende directamente del modelo que aquí se genere.
4. Finalmente, se crea la siguiente generación  $D_k$  de tamaño  $m$  simulando del modelo probabilístico  $M$  que se generó de los mejores individuos.

Los pasos 2, 3 y 4 se repiten hasta alcanzar un criterio de paro previamente establecido, por ejemplo: un número máximo de evaluaciones de función, un número máximo de generaciones o incluso cuando el valor del mejor individuo no cambie en cierto número de generaciones o su cambio sea poco significativo.

El principal problema que se presenta es la estimación del modelo  $M$ , ya que cuanto más complejo sea, mejor recogerá las dependencias entre las variables, pero más compleja/costosa será su estimación.

En la práctica, al implementar un EDA se aplica el *elitismo*, que consiste en conservar al menos uno de los mejores individuos de una generación, para así asegurar que en la siguiente generación no se pierda al mejor individuo.

## 3.2. Clasificación de los EDAs

Los EDAs, en general, pueden trabajar con variables discretas o continuas según los intereses del investigador, en este trabajo, enfocamos nuestra atención en variables continuas, es decir, suponemos que el espacio de búsqueda así como los valores que toman las variables del problema son reales,  $x = (x_1, \dots, x_n) \in \mathbb{R}$ , así que solo trabajaremos con EDAs reales.

Un criterio para la clasificación de los EDAs es el número máximo de dependencias entre las variables (número máximo de padres que cualquier variable puede tener en el modelo gráfico probabilístico) [21].

**Algoritmo 1** Pseudocódigo para EDAs

**Entrada.** Tamaño de la población ( $m$ ), Tamaño de la muestra ( $n$ ). Dimensión ( $d$ ).

```

1:  $D_0 \leftarrow$  Poblacion_inicial ( $m, d$ ).
2:  $k=1$ ;
3:  $f_{apt} \leftarrow g(D_{k-1})$ 
4:  $D_{k-1}^* \leftarrow$  Seleccion( $D_{k-1}, n, f_{apt}$ )
5: mientras no se cumpla condición de paro hacer
6:    $M \leftarrow$  Modelar( $D_{k-1}^*$ )
7:    $D'_{k-1} \leftarrow$  Muestrear ( $M, m$ )
8:    $f_{apt}^* \leftarrow g(D_{k-1})$ 
9:    $f_{apt} \leftarrow f_{apt} \cup f_{apt}^*$ 
10:   $\bar{D}_{k-1} \leftarrow D_{k-1}^* \cup D'_{k-1}$ 
11:   $D_k^* \leftarrow$  Seleccion( $\bar{D}_{k-1}, n, f_{apt}$ )
12:   $k \leftarrow k + 1$ 
13: fin mientras

```

- **Sin dependencias.** Este es uno de los modelos más simples ya que suponen independencia entre todas las variables, en este caso, se asume que la distribución de probabilidad  $n$ -dimensional conjunta se puede factorizar como un producto de  $n$  distribuciones de probabilidad univariadas e independientes. Su modelo de probabilidad sería de la forma:

$$p(x) = \prod_{i=1}^k p_i(x_{j(i)})$$

En muchas ocasiones se trabaja con una distribución normal  $n$ -dimensional, y así es factorizada como un producto de  $n$  distribuciones normales unidimensionales e independientes. Como ejemplos de este grupo se tienen el UMDA<sub>c</sub> (Univariate Marginal Distribution Algorithm), PBIL<sub>c</sub> (Population Based Incremental Learning), CGA<sub>c</sub> (Compact Genetic Algorithm). La letra  $c$  como sub-índice nos indica que es el caso continuo, ya que existe una versión para el caso en el que las variables son discretas. En la Figura 3.2 se muestra la representación gráfica de este tipo de EDAs.

En la práctica, la hipótesis de independencia está muy alejada de lo que realmente ocurre en el problema, en el que generalmente si existen interdependencias entre las variables, así que veamos la siguiente categoría.

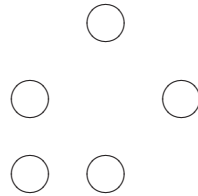


Figura 3.2: Estructura para EDAs con variables independientes.

- **Dependencias Bivariadas.** En este tipo de estructuras se asumen dependencias entre pares de variables para las que se estima una función de densidad bivariada. El modelo de probabilidad es de la forma,

$$p(x) = \prod_{i=1}^k p_i(x_{j_1(i)}, x_{j_2(i)})$$

Como ejemplos de esta categoría se tiene el MIMIC<sub>c</sub> (Mutual Information Maximization for Input Clustering), árbol de dependencias, BM-DA (Bivariate Marginal Distribution Algorithm). En la Figura 3.3 se muestra la representación gráfica de este tipo de EDAs.

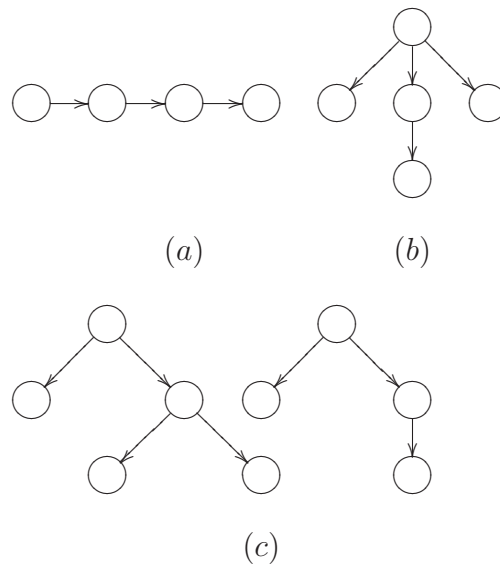


Figura 3.3: Estructuras con dependencias bivariadas: (a) MIMIC, (b) Árbol de dependencias, (c) BMDA

- **Dependencias múltiples.** Este tipo de EDAs son los más complejos, ya que se pueden dar relaciones de dependencia entre dos o más variables.

$$p(x) = \prod_{i=1}^k \psi_i(X_{S_i})$$

. Como ejemplos se tiene el FDA (Factorized Distribution Algorithm), EBNA (Estimation of Bayesian Network Algorithm), BOA (Bayesian Optimization Algorithm), LFDA (Learning Factorized Distribution Algorithm), ECGA (Extended Compact Genetic Algorithm). En la Figura 3.4 se presentan algunas de las estructuras que se pueden generar para los EDAs con dependencias múltiples.



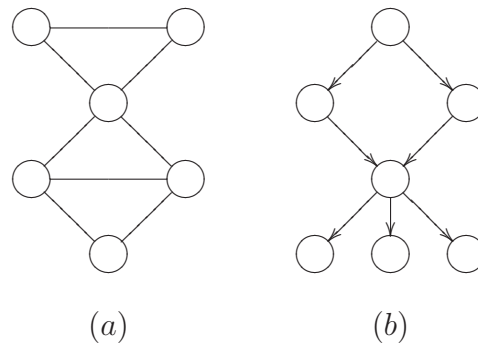


Figura 3.4: Estructuras con dependencias múltiples: (a) FDA, (b) EBNA, BOA

Se debe mencionar que en el presente trabajo elegimos un modelo de dependencias bivariadas, en particular el  $\text{MIMIC}_c$ , ya que como se mencionó anteriormente nos permite estimar una función de distribución bivariada lo que favorece a la propuesta de utilizar cópulas bivariadas como la función de distribución entre cada par de variables, por ello en la siguiente sección conoceremos un poco más acerca del MIMIC.

### 3.3. MIMIC

En [7] se presenta un algoritmo llamado MIMIC (Mutual - Information - Maximizing Input Clustering). Como se menciona anteriormente, el MIMIC es un EDA que trabaja con dependencias bivariadas y como podemos ver en la Figura 3.3 (a) el objetivo es enlazar a las variables una tras otra para formar una cadena, el problema es encontrar el mejor orden de las variables en dicha cadena ya que en un problema de dimensión  $n$  se pueden tener hasta  $n!$  diferentes posibles configuraciones. El modelo probabilístico del MIMIC dada una permutación  $\pi$  es de la forma:

$$p(x) = \prod_{j=1}^{n-1} p(x_{i_{n-j}} | x_{i_{n-j+1}}) \cdot p(x_{i_n}),$$

donde la permutación  $\pi$  es representada por  $(i_1, i_2, \dots, i_n)$ .

Ahora la pregunta sería, ¿Cuál es la mejor permutación para crear la cadena? para ello es necesario definir un criterio que nos permita elegir de

entre las  $n!$  cadenas la que mejor capture las interdependencias entre las variables.

En cada generación se desea encontrar la mejor permutación  $\pi$  tal que su distribución de probabilidad  $p^\pi(x)$  sea la más cercana en divergencia Kullback-Leibler a la distribución empírica del conjunto de individuos seleccionados. Se puede mostrar que la divergencia de Kullback-Leibler entre dos distribuciones de probabilidad,  $p(x)$  y  $p^\pi(x)$  se puede expresar como:

$$H^\pi(x) = h(X_{i_n}) + \sum_{j=1}^{n-1} h(X_{i_j}|X_{i_{j+1}})$$

donde  $h(X)$ , es la entropía de la variable  $X$  y  $h(X|Y)$  es la entropía condicional. Entonces, el problema de buscar la mejor  $p^\pi(x)$  es equivalente a la búsqueda de la permutación  $\pi^*$  que minimiza  $H^\pi(x)$ . Para encontrar  $\pi^*$  en [7] proponen un algoritmo que consiste en seleccionar  $X_{i_n}$  como la variable con menor entropía estimada, posteriormente seleccionar la variable, de entre las variables que aún no han sido seleccionadas, cuya entropía condicionada a la variable seleccionada en el paso anterior sea mínima. En la Figura 2 se muestra el pseudocódigo para la estimación de la distribución de probabilidad conjunta efectuada por el algoritmo MIMIC.

---

**Algoritmo 2** Pseudocódigo para la estimación de la distribución de probabilidad conjunta para el MIMIC

---

**Entrada** Número de variables ( $n$ )

- 1: Seleccionar  $i_n = \operatorname{argmin}_j \hat{h}(X_j)$ .
  - 2: **para**  $k=n-1, \dots, 1$  **hacer**
  - 3:   Escoger  $i_k = \operatorname{argmin}_j \hat{h}(X_j | X_{i_{k+1}}) j \neq i_{k+1}, \dots, i_n$
  - 4: **fin para**
  - 5:  $p^\pi(x) = p(x_{i_1} | x_{i_2}) \cdots p(x_{i_{n-1}} | x_{i_n}) \cdot p(x_{i_n})$
  - 6:  $M = \{i_n, i_{n-1}, \dots, i_1\}$
-

# Capítulo 4

## Funciones de Cópula

Iniciamos este capítulo dando una idea intuitiva sobre el concepto de función de cópula, posteriormente se presentan algunas definiciones necesarias para poder definir una subcópula y a partir de ello poder definir el concepto de cópula.

### 4.1. ¿Qué son las Cópulas?

En estadística, una Cópula es usada como una forma general de formular una distribución multivariada con la característica de que puede representar varios tipos de dependencia. La construcción de una distribución multivariada mediante una cópula se basa en la idea de hacer una transformación simple en cada variable de tal manera que cada variable transformada marginalmente tenga una distribución uniforme. Una vez que se tiene tal transformación marginal, la estructura de dependencia puede ser expresada como una distribución multivariada sobre los uniformes obtenidos y una cópula es precisamente eso, una distribución multivariada sobre variables aleatorias uniformes [33].

En 1959, Abe Sklar utilizó por primera vez la palabra **Cópula** debido a su significado gramatical, palabra o expresión que enlaza un sujeto y predicado, pero existen trabajos de Hoeffding del año de 1940 los que trabaja con esta teoría [3].

Ahora, introduciremos la notación y la definición de algunos conceptos importantes para el estudio de las cópulas [28]. Sea  $\mathbb{R}$  la línea real  $(-\infty, \infty)$ ,  $\bar{\mathbb{R}}$  denota la línea real extendida  $[-\infty, \infty]$  y  $\bar{\mathbb{R}}^2$  denota el plano real extendido

$\bar{\mathbb{R}} \times \bar{\mathbb{R}}$ . Un rectángulo en  $\bar{\mathbb{R}}^2$  es el producto cartesiano  $B$  de dos intervalos cerrados:  $B = [x_1, x_2] \times [y_1, y_2]$ . Los vértices del rectángulo  $B$  son los puntos  $(x_1, y_1)$ ,  $(x_1, y_2)$ ,  $(x_2, y_1)$  y  $(x_2, y_2)$ . El cuadro unitario  $I^2$  es el producto  $I \times I$  donde  $I = [0, 1]$ . Una función real 2-valuada  $H$  es una función cuyo dominio,  $\text{Dom}H$ , es un subconjunto de  $\bar{\mathbb{R}}^2$  y cuyo rango,  $\text{Ran}H$ , es un subconjunto de  $\mathbb{R}$ .

Para introducir el concepto de función de cópula, es necesario presentar las siguientes definiciones:

**Definición 4.1.1. Volumen de un cuadrado  $B$  bajo una función  $H$ .**

Sean  $S_1$  y  $S_2$  dos subconjuntos no vacíos de  $\bar{\mathbb{R}}$ , y sea  $H$  una función real tal que  $\text{Dom}H = S_1 \times S_2$ . Sea  $B = [x_1, x_2] \times [y_1, y_2]$  un rectángulo cuyos vértices están en  $\text{Dom}H$ . Entonces el  $H$ -volumen de  $B$  está dado por

$$V_H(B) = H(x_2, y_2) - H(x_2, y_1) - H(x_1, y_2) + H(x_1, y_1) \quad (4.1)$$

**Definición 4.1.2. Función  $H$  creciente** Una función real 2-valuada  $H$  es 2-creciente si  $V_H(B) \geq 0$  para todos los rectángulos  $B$  cuyos vértices están en  $\text{Dom}H$ .

**Lema 2.** Sean  $S_1$  y  $S_2$  dos subconjuntos no vacíos de  $\bar{\mathbb{R}}$ , y sea  $H$  una función 2-creciente con dominio  $S_1 \times S_2$ . Sean  $x_1, x_2$  en  $S_1$  con  $x_1 \leq x_2$  y  $y_1, y_2$  en  $S_2$  con  $y_1 \leq y_2$ . Entonces la función  $t \mapsto H(t, y_2) - H(t, y_1)$  es no decreciente sobre  $S_1$ , y la función  $t \mapsto H(x_2, t) - H(x_1, t)$  es no decreciente sobre  $S_2$ .

**Definición 4.1.3. Función  $H$  aterrizada.** Sea  $H$  definida como antes y supongamos que  $S_1$  tiene elemento mínimo  $a_1$  y  $S_2$  elemento mínimo  $a_2$ . Decimos que  $H$  es aterrizada si para todo  $(x, y) \in S_1 \times S_2$  se cumple  $H(x, a_2) = H(a_1, y) = 0$ .

**Lema 3.** Sea  $H$  una función aterrizada, 2-creciente. Entonces  $H$  es no decreciente en cada argumento.

Con las definiciones presentadas, ya podemos introducir el concepto de función de cópula, pero antes es necesario definir primero una subcópula,

**Definición 4.1.4.** Una subcópula 2 – dimensional (o 2-subcópula) es una función  $C'$  con las siguientes propiedades:

1.  $\text{Dom}C' = S_1 \times S_2$ , donde  $S_1$  y  $S_2$  son subconjuntos de  $\mathbb{I}$

2.  $C'$  es una función aterrizada y 2-creciente.

3. Para cada  $u$  en  $S_1$  y para cada  $v$  en  $S_2$ ,

$$C'(u, 1) = u \text{ y } C'(1, v) = v$$

**Definición 4.1.5.** Una **cópula**  $C$  2 – dimensional (o 2-cópula) es una 2-subcópula  $C$  cuyo dominio es  $\mathbb{I}^2$ . De manera equivalente, una cópula es una función aterrizada, 2 – creciente cuyo dominio es el cuadro unitario  $I^2$ , que satisface,

1. Para cada  $u, v$  en  $I$

$$C(u, 0) = 0 = C(0, v)$$

y

$$C(u, 1) = u, \quad C(1, v) = v;$$

2. Para cada  $u_1, u_2, v_1, v_2$  en  $I$  tal que  $u_1 \leq u_2$  y  $v_1 \leq v_2$ ,

$$C(u_2, v_2) - C(u_2, v_1) - C(u_1, v_2) + C(u_1, v_1) \geq 0.$$

La diferencia entre una cópula y una subcópula es el dominio, sin embargo muchas propiedades importantes de subcópulas son también propiedades de cópulas, veamos el siguiente teorema.

**Teorema 1.** [6] Sea  $C'$  una subcópula. Entonces para cada  $(u, v)$  en  $DomC'$ ,

$$\max(u + v - 1, 0) \leq C'(u, v) \leq \min(u, v).$$

Dado que una cópula es una subcópula la desigualdad del Teorema 1 también se cumple para cópulas.

Las cotas del teorema anterior son conocidas como “Cotas de Fréchet-Hoeffding ” y las denotaremos como sigue:

◦ En la Figura 4.1 se tiene la cota inferior de Fréchet-Hoeffding, la cual se define de la siguiente manera:

$$W(u, v) = \max(u + v - 1, 0).$$

- En la Figura 4.2 se presenta la cota superior de Fréchet-Hoeffding que esta definida como:

$$M(u, v) = \min(u, v).$$

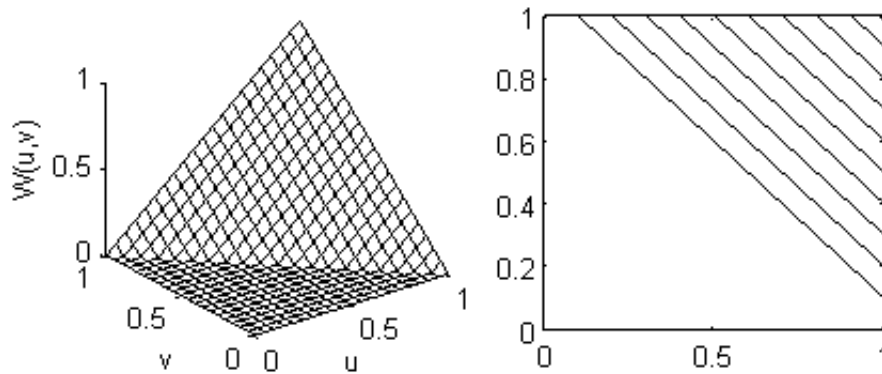


Figura 4.1: Cota inferior de Fréchet-Hoeffding (izquierda) y sus curvas de nivel (derecha)

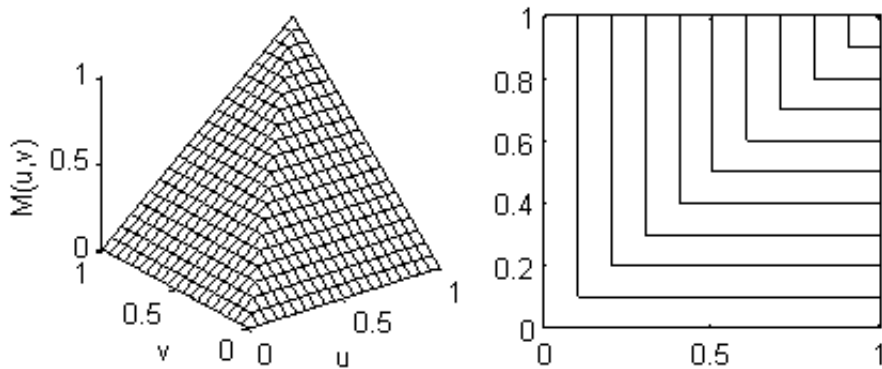


Figura 4.2: (Izquierda) Cota Superior de Fréchet-Hoeffding, (Derecha) Curvas de nivel de la cota superior

En dimensión 2, las cotas de Fréchet-Hoeffding también son cópulas, pero no es así para dimensiones mayores que 2, aunque la  $n$ -cópula sigue estando

acotada. Como se ha mencionado, la función de cópula permite expresar la función de distribución conjunta de las variables en términos de su distribución marginal univariada y ese resultado queda expresado en el siguiente teorema.

## 4.2. Teorema de Sklar

Un resultado importante en la teoría de Cópulas es el Teorema de Sklar (1959), ya que garantiza la relación de una función de distribución conjunta con sus distribuciones marginales.

**Teorema 2.** (*Teorema de Sklar*) Sea  $H$  una función de distribución conjunta con marginales  $F$  y  $G$ . Existe una cópula  $C$  tal que para todo  $x, y$  en  $\mathbb{R}$ ,

$$H(x, y) = C(F(x), G(y)).$$

Si  $F$  y  $G$  son funciones continuas, entonces  $C$  es única; en caso contrario,  $C$  está únicamente determinada en  $\text{Ran}F \times \text{Ran}G$ . Recíprocamente, si  $C$  es una cópula y  $F$  y  $G$  son funciones de distribución, entonces la función  $H$  definida por  $H(x, y) = C(F(x), G(y))$  es una función de distribución conjunta con marginales  $F$  y  $G$ .

Si se desea consultar sobre la Teoría de cópulas se recomienda consultar [28, 30].

### 4.2.1. Función de densidad de una cópula

Las cópulas, al igual que las funciones de distribución de variables aleatorias continuas, admiten la noción de función de densidad, veamos la siguiente definición:

**Definición 4.2.1.** La función de densidad  $c(u, v)$  asociada a la cópula  $C(u, v)$  esta dada por,

$$c(u, v) = \frac{\partial^2 C(u, v)}{\partial u \partial v}$$

Para variables aleatorias continuas, la densidad de la cópula esta relacionada con la densidad de la función de distribución  $F_{XY}$ , denotada como  $f_{XY}$ , veamos la siguiente definición:

**Definición 4.2.2.** Si  $f_{XY}$  y  $f_X, f_Y$  denotan las funciones de densidad conjunta y las densidades marginales de las variables aleatorias  $X, Y$  respectivamente, con función de distribución conjunta  $F_{XY}$ , entonces

$$c(F_X(x), F_Y(y)) = \frac{f_{XY}(x, y)}{f_X(x) f_Y(y)} \quad (4.2)$$

De (4.2) se tiene la **representación canónica de una cópula**, que es de la siguiente forma:

$$f_{XY}(x, y) = c(F_X(x), F_Y(y)) f_X(x) f_Y(y).$$

### 4.2.2. Algunas Cópulas comunes.

Existen muchas familias de cópulas, uno de los criterios para su clasificación es su número de parámetros de dependencia. Nosotros sólo trabajaremos con cópulas 1-paramétricas. Dadas las definiciones anteriores, es el momento de presentar algunas de las cópulas más estudiadas y, por consecuencia, las más utilizadas. Presentaremos la función de la cópula así como su función de densidad [30].

1. Cópula producto.

Es la cópula más simple, ya que no tiene parámetro de dependencia. Su función de distribución esta dada como sigue:

$$C(u, v) = uv,$$

donde,  $u$  y  $v$  toman valores del intervalo unitario. Y su función de densidad está dada por

$$c(u, v) = 1,$$

ésta cópula es importante ya que representa independencia en los datos.

2. Cópula Ali-Mikhail-Haq (AMH).

Su función de distribución es:

$$C(u, v) = uv(1 - \theta(1 - u)(1 - v))^{-1}$$



La función de densidad es:

$$c(u, v) = \frac{1 + \theta(u + v + uv - 2) - \theta^2(u + v - uv - 1)}{(1 - \theta(1 - u)(1 - v))^3}.$$

Su parámetro de dependencia puede tomar valores en  $[-1, 1]$ , pero modela dependencias débiles. En la Figura 4.4, se presenta tanto la función de distribución como la función de densidad de la cópula AMH con parámetro de dependencia  $\theta = -0.5$ .

### 3. Cópula Clayton.

Originalmente fue estudiada por Kimeldorf y Sampson en 1975 [18], tomando la definición de Nelsen (1999), su función de distribución es,

$$C(u, v; \theta) = \max\left\{(u^{-\theta} + v^{-\theta} - 1)^{-1/\theta}, 0\right\},$$

su función de densidad es:

$$c(u, v; \theta) = (1 + \theta)(uv)^{-(\theta+1)} \left(u^{-\theta} + v^{-\theta} - 1\right)^{-2-1/\theta}$$

su parámetro de dependencia puede tomar valores de  $(-1, \infty) \setminus \{0\}$ . Cuando el parámetro  $\theta$  tiende a infinito, la cópula Clayton tiende a la cópula límite superior de Fréchet, un inconveniente de la cópula Clayton es que no puede capturar dependencias negativas. En la Figura 4.5, se presenta la función de distribución y la función de densidad de la cópula Clayton con parámetro de dependencia  $\theta = 2$ .

### 4. Cópula de Farlie-Gumbel-Morgenstern (FGM).

Originalmente fue propuesta por Morgenstern en 1956 [25], su función de distribución es:

$$C(u, v) = uv(1 + \theta(1 - u)(1 - v)).$$

La función de densidad es:

$$c(u, v) = 1 + \theta(1 - 2u)(1 - 2v).$$

La cópula FGM es una perturbación de la cópula producto; su parámetro de dependencia,  $\theta$ , toma valores en  $[-1, 1]$ , si  $\theta$  es igual a cero, entonces

la cópula FGM converge a la cópula independiente. La limitante de esta cópula es que solo se puede usar cuando la dependencia entre las marginales acumuladas de las variables es modesta en magnitud. En la Figura 4.5, se presenta la función de distribución y la función de densidad de la cópula FGM con parámetro de dependencia  $\theta = -0.8$ .

#### 5. Cópula Frank.

Su función de distribución es:

$$C(u, v; \theta) = -\frac{1}{\theta} \log \left( 1 + \frac{(e^{-\theta u} - 1)(e^{-\theta v} - 1)}{e^{-\theta} - 1} \right)$$

la función de densidad es:

$$c(u, v; \theta) = \frac{-\theta(e^{-\theta} - 1)e^{-\theta(u+v)}}{((e^{-\theta} - 1) + (e^{-\theta u} - 1)(e^{-\theta v} - 1))^2}.$$

El parámetro de dependencia puede tomar cualquier valor real de  $(-\infty, \infty) \setminus \{0\}$ . Los valores de  $-\infty, 0, \infty$  corresponden al límite inferior de Fréchet, a la cópula independiente y al límite superior de Fréchet, respectivamente. En la Figura 4.6, se presenta la función de distribución y la función de densidad de la cópula Frank con parámetro de dependencia  $\theta = 8$ .

#### 6. Cópula Gaussiana (Normal).

Fue propuesta por Lee en 1983, su función de distribución es de la forma

$$C(u, v; \theta) = \Phi_G(\Phi^{-1}(u), \Phi^{-1}(v)),$$

donde  $\Phi$  es la función de distribución de la normal estándar, y  $\Phi_G(u, v)$  es la función de distribución normal estándar bivariada cuyo parámetro de correlación es  $\theta$  sobre el intervalo  $(-1, 1)$ .

La función de densidad de la cópula gaussiana es:

$$c(u_i, v_i; \theta) = \frac{-1}{\sqrt{1 - \theta^2}} \exp \left\{ -\frac{x_i^2 + y_i^2 - 2\theta x_i y_i}{2(1 - \theta^2)} \right\} \exp \left\{ -\frac{1}{2}(x_i^2 + y_i^2) \right\}$$

donde,  $x_i = \Phi^{-1}(u_i)$  y  $y_i = \Phi^{-1}(v_i)$ . La cópula normal es flexible, en el sentido de que permite dependencias tanto positivas como negativas e incluso independencia. En la Figura 4.7, se presenta la función de distribución y la función de densidad de la cópula Gaussiana con parámetro de dependencia  $\theta = 0.8$ .

7. Cópula Gumbel.

Es de la forma,

$$C(u, v; \theta) = \exp\left(-(\tilde{u}^\theta + \tilde{v}^\theta)^{1/\theta}\right),$$

su función de densidad es:

$$c(u, v; \theta) = \frac{C(u, v)}{uv} \frac{(\tilde{u}\tilde{v})^{\theta-1}}{(\tilde{u}^\theta + \tilde{v}^\theta)^{2-1/\theta}} \left( ((\tilde{u}^\theta + \tilde{v}^\theta)^{1/\theta} + \theta - 1) \right)$$

donde,  $\tilde{u} = -\log(u)$  y  $\tilde{v} = -\log(v)$ . El parámetro de dependencia se restringe al intervalo  $[1, \infty)$ . Los valores de 1 y  $\infty$  corresponden a la cópula independiente y al límite superior de Fréchet, respectivamente. Al igual que la cópula Clayton, ésta cópula no modela dependencias negativas. En la Figura 4.8, se presenta la función de distribución y la función de densidad de la cópula Gumbel con parámetro de dependencia  $\theta = 2$ .

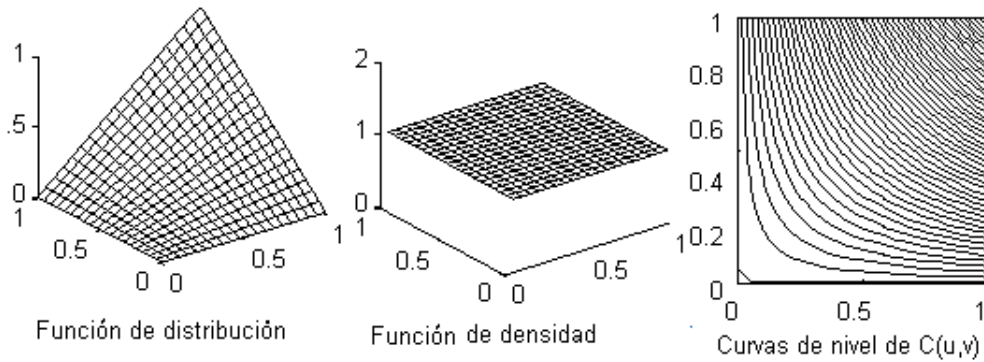


Figura 4.3: Cópula Producto

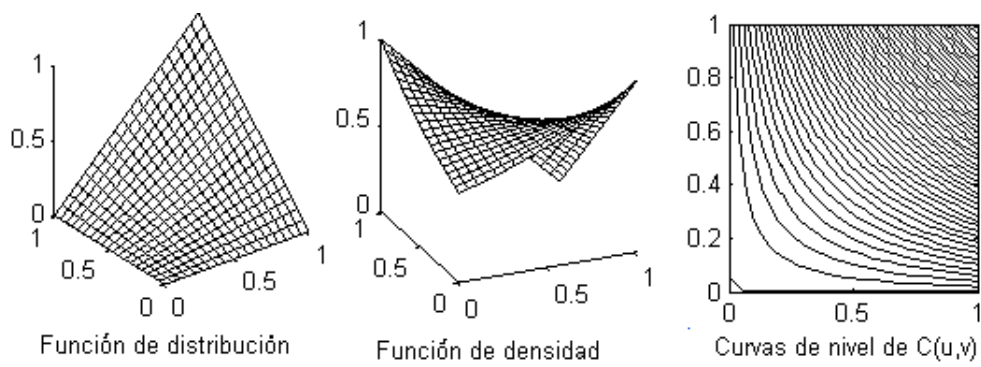


Figura 4.4: Cópula de Ali-Mikhail-Haq, con parámetro  $\theta = -0.5$

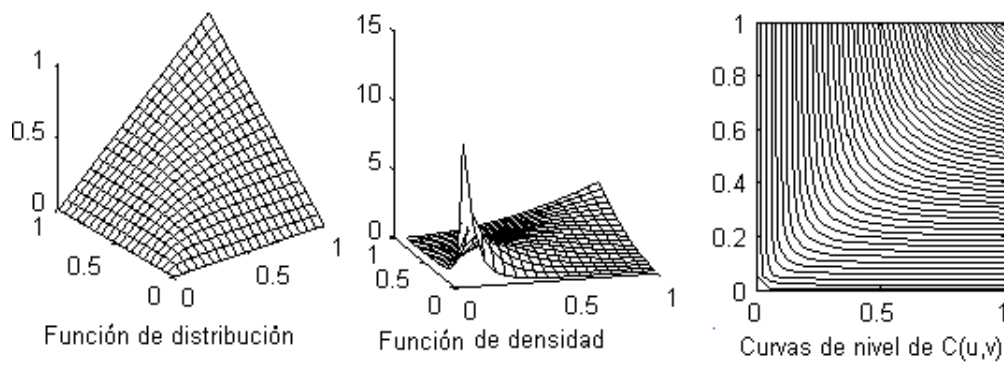


Figura 4.5: Cópula Clayton, con parámetro  $\theta = 2$

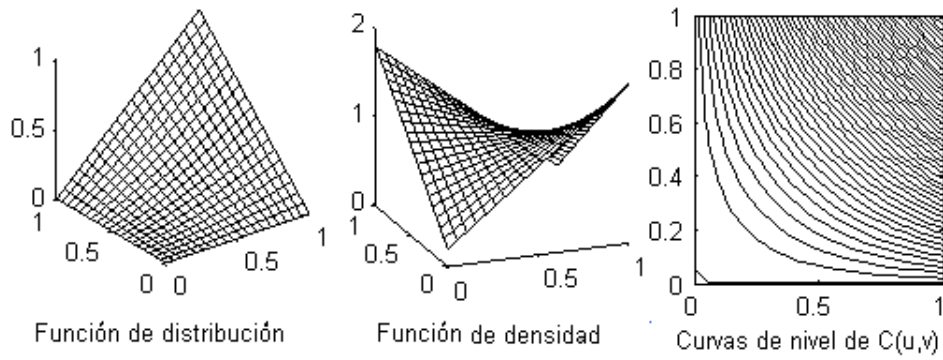


Figura 4.6: Cópula de Farlie-Gumbel-Morgenstern, con parámetro  $\theta = -0.8$

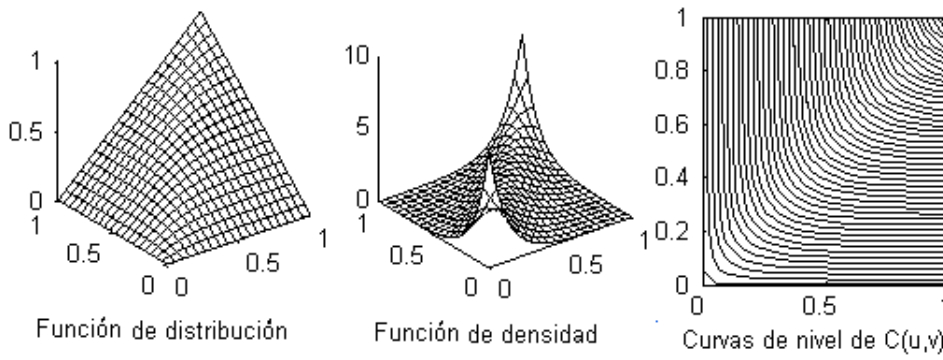
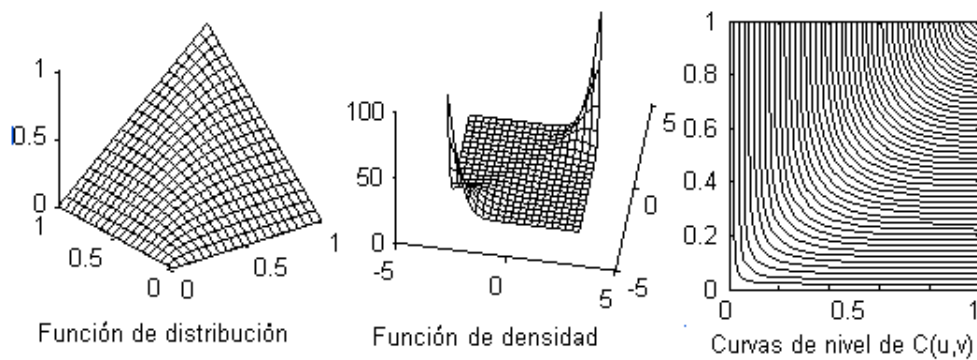
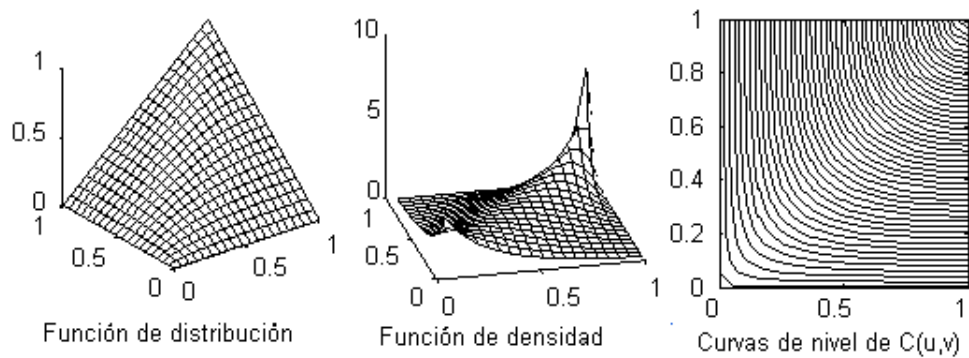


Figura 4.7: Cópula Frank, con parámetro  $\theta = 8$

Figura 4.8: Cópula Gaussiana, con parámetro  $\theta = 0.8$ Figura 4.9: Cópula Gumbel, con parámetro  $\theta = 2$ 

Se han seleccionado las cópulas 1-paramétricas ya que la intención es poder capturar las interrelaciones entre variables, y aquí haremos uso de algunas definiciones presentadas en el Capítulo 2 sobre medidas de dependencia.

### 4.2.3. Medidas de dependencia para cópulas

Como hemos visto, las cópulas son funciones que relacionan la distribución conjunta de dos variables con sus distribuciones marginales. El objetivo principal de trabajar los EDAs con funciones de cópula es poder capturar y

mantener las dependencias entre variables. Ahora se presentan las medidas de dependencia en términos de cópulas ya que las podemos interpretar como funciones de probabilidad.

La medida de correlación por rangos  $\rho_r$  de Spearman y  $\tau$  de Kendall se pueden expresar en términos de cópulas (Nelsen, 1999), veamos la siguiente proposición.

**Proposición 4.2.1.** [17] Si  $X$  y  $Y$  son variables aleatorias continuas con función de distribución conjunta  $F_{XY}$  y distribuciones marginales  $F_X$  y  $F_Y$  respectivamente, sea  $C$  una cópula, tal que,  $F_{XY}(x, y) = C(F_X(x), F_Y(y))$ , entonces  $\tau$  de Kendall y  $\rho_r$  de Spearman para  $X$  y  $Y$  pueden ser expresados como:

$$\begin{aligned}\tau &= 4 \int_{[0,1]^2} C(s, t) dC(s, t) - 1, \\ \rho_r &= 12 \int_{[0,1]^2} st dC(s, t) - 3 = 12 \int_{[0,1]^2} C(s, t) ds dt - 3\end{aligned}$$

En la Tabla 4.1 se presentan las medidas de correlación por rangos para algunas cópulas, en ella se puede observar que tanto  $\rho_r$  como  $\tau$ , quedan expresadas en función del parámetro de dependencia de la cópula. Entonces, dada una muestra aleatoria, es relativamente fácil encontrar un valor para el parámetro  $\theta$ , ya que se puede estimar el coeficiente de correlación por rangos ( $\rho_r$  ó  $\tau$ ) y resolver la ecuación que corresponda a la cópula con la que se trabaja y así encontrar un valor para el parámetro de dependencia. Otro método para estimar el parámetro  $\theta$  es mediante máxima verosimilitud y sobre eso se habla en el siguiente capítulo.

Cópula	Dominio del parámetro	$\tau$ de Kendall	$\rho_r$ de Spearman
AMH	$-1 \leq \theta \leq 1$	$\left(\frac{3\theta-2}{3\theta}\right) - \frac{2}{3}\left(1 - \frac{1}{\theta}\right)^2 \ln(1 - \theta)$	$\frac{1}{3}\theta$
Clayton	$\theta \in [-1, \infty) \setminus \{0\}$	$\frac{\theta}{\theta+2}$	*
FGM	$[-1, 1]$	$\frac{2}{9}\theta$	$\frac{1}{3}\theta$
Frank	$(-\infty, \infty)$	$1 - \frac{4}{\theta}[1 - D_1(\theta)]$	$1 - \frac{12}{\theta}[D_1(\theta) - D_1 - 2(\theta)]$
Gaussiana	$-1 < \theta < 1$	$\frac{2}{\pi} \arcsin(\theta)$	$\frac{6}{\pi} \arcsin\left(\frac{\theta}{2}\right)$

Tabla 4.1: Expresiones para la correlación en términos del parámetro de dependencia de la cópula.

\* No se tiene una expresión explícita para  $\rho_r$  en términos de  $\theta$ .

$D_1(x)$  es la función Debay y está dada como  $D_1(x) = \frac{1}{x} \int_0^x \frac{t}{e^t-1} dt$



# Capítulo 5

## Inferencia Estadística para cópulas

Debemos notar que todas las cópulas presentadas, a excepción de la cópula producto, dependen de un parámetro  $\theta$ , llamado **parámetro de dependencia**. Al trabajar con cópulas, nos debemos enfrentar con dos problemas principales, el primero de ellos consiste en elegir la cópula que mejor capture las dependencias entre las variables, el segundo se presenta cuando ya se eligió de alguna manera la cópula con la que se desea trabajar, poder estimar su parámetro de dependencia. Para comenzar esta sección iniciaremos presentando algunos métodos para estimar el parámetro de dependencia de la cópula y posteriormente se presenta la propuesta de Ané y Kharoubi (2003) [2] para la selección de cópulas que se basa en la construcción de la cópula empírica.

Existen diferentes procedimientos para estimar los coeficientes de un modelo de regresión, o para estimar los parámetros de una distribución de probabilidad. De entre esos procedimientos probablemente el más versátil, ya que se puede aplicar en gran cantidad de situaciones, y por ello uno de los más empleados es el método de máxima verosimilitud.

### 5.1. Método Exacto de Máxima Verosimilitud

Desde el punto de vista estadístico, una cópula no es más que un modelo multivariado, en nuestro caso es un modelo bivariado, y con la finalidad de

estimar su parámetro de dependencia haremos uso de la teoría sobre Estimadores de máxima verosimilitud (Maximun Likelihood Estimation, EML).

Daniel Bernoulli en 1777 [4] fue el primero en aplicar la idea de verosimilitud para la solución de un problema de estimación. Posteriormente, R. A. Fisher (1921) [12] retoma, generaliza y formaliza la idea aunque ya de una forma mucho más clara y práctica.

Supongamos que se tiene un modelo de probabilidad para un experimento, el cual involucra un parámetro desconocido  $\theta$ . Se realiza el experimento y se obtienen algunos datos. Lo que se desea es utilizar los datos para estimar el parámetro  $\theta$ . Los datos observados se pueden considerar como un evento  $E$  en el espacio muestral del modelo de probabilidad. Entonces la probabilidad del evento  $E$  puede quedar determinada como  $P(E; \theta)$ .

Luego, el estimador de máxima verosimilitud de  $\theta$  ( $\hat{\theta}$ ) es el valor de  $\theta$  que maximiza  $P(E; \theta)$  [17]. Entonces, la función de verosimilitud de  $\theta$  se define como sigue:

$$L(\theta) = k \cdot P(E; \theta)$$

donde  $k$  es cualquier constante respecto a  $\theta$ .

Usualmente  $P(E; \theta)$  y  $L(\theta)$  son productos de algunos términos, por lo que conviene trabajar con logaritmos, por lo que definimos la función *log – verosimilitud* como:

$$l(\theta) = \log(L(\theta))$$

Entonces,

$$l(\theta) = \log(k) + \log(P(E; \theta))$$

Como  $k$  no depende de  $\theta$ , el estimador  $\hat{\theta}$  es el valor de  $\theta$  que maximiza  $P(E; \theta)$ , también maximiza  $L(\theta)$  y  $l(\theta)$ .

Se ha dicho que  $\hat{\theta}$  es el valor que maximiza una función de  $\theta$ , entonces para hallar ese máximo solo debemos hallar la derivada de la función, igualar a cero y resolver para  $\theta$ , con esa finalidad definimos la función de Score como sigue:

$$S(\theta) = l'(\theta) = \frac{dl(\theta)}{d\theta},$$

y la función de información  $\mathcal{I}(\theta)$  se define como:

$$\mathcal{I}(\theta) = -l''(\theta) = -S'(\theta) = -\frac{d^2l(\theta)}{d\theta^2}.$$

Entonces, para garantizar que  $\hat{\theta}$  es un valor de  $\theta$  que maximiza a  $L(\theta)$  se debe cumplir que,

$$S(\hat{\theta}) = 0; \quad \mathcal{I}(\hat{\theta}) > 0.$$

Ahora, se muestra el desarrollo para estimar el parámetro de dependencia de una cópula mediante máxima verosimilitud.

En principio supondremos que  $f(x; \theta)$  es cualquier función de densidad,  $\theta$  el parámetro (o conjunto de parámetros) que deseamos estimar, y  $x_i$  una muestra aleatoria e independiente del experimento, de tamaño  $N$ , entonces la probabilidad de las  $N$  observaciones  $x_1, x_2, \dots, x_N$  es:

$$P(x_1, x_2, \dots, x_N; \theta) = f(x_1; \theta)f(x_2; \theta)\dots f(x_N; \theta)$$

Luego,

$$\begin{aligned} L(\theta; x_i) &= kP(x_1, x_2, \dots, x_N; \theta) \\ L(\theta; x_i) &= kf(x_1; \theta)f(x_2; \theta)\dots f(x_N; \theta) \\ L(\theta; x_i) &\propto \prod_{i=1}^N f(x_i; \theta) \end{aligned} \quad (5.1)$$

Obtenemos la función *log – verosimilitud*,  $l(\theta)$  como:

$$\begin{aligned} l(\theta; x_i) &= \log(L(\theta; x_i)) \\ l(\theta; x_i) &= \log\left(\prod_{i=1}^N f(x_i; \theta)\right) \\ l(\theta; x_i) &= \sum_{i=1}^N \log(f(x_i; \theta)) \end{aligned} \quad (5.2)$$

por último buscamos maximizar la función  $l(\theta; x_i)$ .

Recordemos la representación canónica de una cópula,

$$f(x_1, x_2, \dots, x_N) = c(F_1(x_1), F_2(x_2), \dots, F_n(x_n)) \cdot \prod_{j=1}^n f_j(x_j) \quad (5.3)$$

donde

$$c(F_1(x_1), F_2(x_2), \dots, F_n(x_n)) = \frac{\partial^n (C(F_1(x_1), F_2(x_2), \dots, F_n(x_n)))}{\partial F_1(x_1) \partial F_2(x_2) \dots \partial F_n(x_n)}$$

es la  $n$ -ésima derivada parcial de la cópula  $C$ , es decir,  $c$  es la densidad de la cópula y  $f_j$  es la  $j$ -ésima función de densidad univariada.

En particular, si la función de densidad  $f(x; \theta)$  en (5.2) es de la forma (5.3) y  $\theta$  es el conjunto (vector) de parámetros de la cópula y de las funciones marginales, entonces la función *log-likelihood* que se debe maximizar es de la siguiente forma:

$$\begin{aligned} l(\theta; x_i) &= \sum_{i=1}^N \log \left( c(F_1(x_{1i}), F_2(x_{2i}), \dots, F_n(x_{ni})) \cdot \prod_{j=1}^n f_j(x_{ji}) \right) \\ &= \sum_{i=1}^N \log(c(F_1(x_{1i}), F_2(x_{2i}), \dots, F_n(x_{ni}))) \\ &\quad + \sum_{i=1}^N \sum_{j=1}^n \log f_j(x_{ji}) \end{aligned} \quad (5.4)$$

La complejidad de éste procedimiento radica en que se debe maximizar una función  $(n + n_c)$ -dimensional, donde  $n$  es la dimension de la cópula y  $n_c$  es el número de parámetros de la cópula, (recordemos que en este trabajo solo se utilizan cópulas 2-dimensionales y 1-paramétricas), dado el inconveniente, se presentan dos alternativas para atacar el problema.

## 5.2. Método IFM

El método de máxima verosimilitud puede ser muy caro computacionalmente cuando se trabaja en altas dimensiones, por ello en 1996 Joe y Xu [16] propusieron una alternativa para la estimación del conjunto de parámetros  $\theta$ .

Si observamos la Ecuación (5.4) podemos notar que es la suma de dos expresiones, la primera de ellas involucra los parámetro de la cópula y los parámetros de las marginales, pero el segundo término sólo depende de los parámetros de las funciones marginales, así que la propuesta es maximizar cada una de las partes por separado. El método IFM (Inference For the Margins) consta de dos pasos:

1. El primer paso consiste en estimar los parámetros de las marginales  $\theta_1$

$$\hat{\theta}_1 = \text{ArgMax}_{\theta_1} \sum_{i=1}^N \sum_{j=1}^n \log f_j(x_{ji}; \theta_1)$$

2. El segundo paso consiste en que dado el vector de parámetros  $\hat{\theta}_1$  se estime el parámetro de la cópula  $\theta_2$ :

$$\hat{\theta}_2 = \underset{\theta_2}{\text{ArgMax}} \sum_{i=1}^N \log(c(F_1(x_{1i}), F_2(x_{2i}), \dots, F_n(x_{ni}); \theta_2, \hat{\theta}_1))$$

Entonces el estimador IFM se define como:

$$\hat{\theta}_{IFM} = (\hat{\theta}_1, \hat{\theta}_2)'$$

Con éste método lo que se logra es bajar la dimensión en el problema de maximizar la verosimilitud, pero notemos que una de las suposiciones que estamos haciendo, es que la función de densidad de la variable  $x_i$  es  $f_i(x_i)$  y más aún, que su función de distribución es  $F_i$ , lo cual no se cumple en nuestro caso, ya que no sabemos que distribución siguen las variables con las que trabajamos, por ello presentamos el siguiente método de estimación, en el que no será necesario conocer la función de distribución de las variables.

### 5.3. Método CML

En los dos métodos de inferencia presentados, se asume que conocemos la distribución marginal que sigue cada una de las variables, sin embargo en muchas ocasiones eso no sucede. Existe un tercer método de inferencia para cópulas llamado Canonical Maximum Likelihood (CML), consiste en transformar la muestra  $\{x_{1i}, x_{2i}, \dots, x_{ni}\}_{i=1}^N$  en variables uniformes  $\{u_{1i}, u_{2i}, \dots, u_{ni}\}_{i=1}^N$  y así solamente estimar el parámetro de la cópula. El método CML se puede describir como sigue:

1. Realizar la transformación de las variables  $\{x_{1i}, x_{2i}, \dots, x_{ni}\}_{i=1}^N$  a variables uniformes mediante su distribución marginal empírica, es decir, estimar  $\hat{F}_i(x_{ij})$  para  $i = 1, \dots, n$ .
2. Estimar mediante MLE el parámetro de la cópula.

$$\hat{\theta}_2 = \underset{\theta_2}{\text{ArgMax}} \sum_{i=1}^N \log(c(\hat{F}_1(x_{1i}), \hat{F}_2(x_{2i}), \dots, \hat{F}_n(x_{ni}); \theta_2))$$

Se han presentado tres métodos para la estimación de parámetros, se debe señalar que el método que utilizamos es el CML, ya que no conocemos a priori la distribución de las variables y así solo debemos estimar un parámetro, el de la cópula.

## 5.4. Ejemplos de estimación de parámetro para cópulas

Veamos algunos ejemplos de la estimación del parámetro de dependencia, con el método CML, para las cópulas que hemos presentado en capítulos anteriores.

Planteamiento: Considere una muestra de variables aleatorias  $X, Y$ . Se desea estimar el parámetro de dependencia para alguna cópula dada.

1. El paso número uno es similar para todas las cópulas ya que consiste en calcular la función de distribución empírica para estimar la distribución teórica de la cópula. Entonces, sea  $X = \{x_1, x_2, \dots, x_N\}$  y  $Y = \{y_1, y_2, \dots, y_N\}$  la muestra de variables aleatorias de tamaño  $N$  y sean  $u = \{\hat{F}_X(X)\}$  y  $v = \{\hat{F}_Y(Y)\}$  las transformaciones de la muestra mediante la distribución marginal empírica, así  $u, v \sim U(0, 1)$ .
2. Dada una cópula, el paso número dos consiste en aplicar al método de máxima verosimilitud, veamos los cálculos que se hacen para cada cópula.

**Cópula Ali-Mikhail-Haq** Del método de máxima verosimilitud se tiene que la función de verosimilitud  $L(\theta)$  es,

$$\begin{aligned} L(\theta) &\propto \prod_{i=1}^N c(u_i, v_i; \theta) \\ &\propto \prod_{i=1}^N \left( \frac{1 + \theta(u_i + v_i + u_i v_i - 2) - \theta^2(u_i + v_i - u_i v_i - 1)}{(1 - \theta(1 - u_i)(1 - v_i))^3} \right) \end{aligned}$$

y la función log-verosimilitud es:

$$\begin{aligned} l(\theta) &= \sum_{i=1}^N \log \left( \frac{1 + \theta(u_i + v_i + u_i v_i - 2) - \theta^2(u_i + v_i - u_i v_i - 1)}{(1 - \theta(1 - u_i)(1 - v_i))^3} \right) \\ &= \sum_{i=1}^N \left( \log [1 + \theta(u_i + v_i + u_i v_i - 2) - \theta^2(u_i + v_i - u_i v_i - 1)] \right. \\ &\quad \left. - 3 \log [1 - \theta(1 - u_i)(1 - v_i)] \right) \end{aligned}$$

Derivando  $l(\theta)$  respecto a  $\theta$  obtenemos la función de Score,

$$S(\theta) = \sum_{i=1}^N \left( \frac{u_i + v_i + u_i v_i - 2 - 2\theta(u_i + v_i - u_i v_i - 1)}{1 + \theta(u_i + v_i + u_i v_i - 2) - \theta^2(u_i + v_i - u_i v_i - 1)} + \frac{3(1 - u_i)(1 - v_i)}{1 - \theta(1 - u_i)(1 - v_i)} \right)$$

y sólo resta resolver  $S(\theta) = 0$  para  $\theta$ .

**Cópula Clayton** Del método de máxima verosimilitud se tiene que la función de verosimilitud  $L(\theta)$  es,

$$\begin{aligned} L(\theta) &\propto \prod_{i=1}^N c(u_i, v_i; \theta) \\ &\propto \prod_{i=1}^N \left( (1 + \theta)(u_i v_i)^{-(\theta+1)} \left( u_i^{-\theta} + v_i^{-\theta} - 1 \right)^{-2-1/\theta} \right) \end{aligned}$$

y la función log-verosimilitud es:

$$\begin{aligned} l(\theta) &= \sum_{i=1}^N \log \left( (1 + \theta)(u_i v_i)^{-(\theta+1)} \left( u_i^{-\theta} + v_i^{-\theta} - 1 \right)^{-2-1/\theta} \right) \\ &= N \log(1 + \theta) - (\theta + 1) \sum_{i=1}^N \log(u_i v_i) \\ &\quad - \left( \frac{1}{\theta} + 2 \right) \sum_{i=1}^N \log(u_i^{-\theta} + v_i^{-\theta} - 1) \end{aligned}$$

Derivando  $l(\theta)$  respecto a  $\theta$  obtenemos la función de Score,

$$\begin{aligned} S(\theta) &= \frac{N}{1 + \theta} - \left[ \left( \frac{1}{\theta} + 2 \right) \sum_{i=1}^N \frac{-u_i^{-\theta} \log(u_i) - v_i^{-\theta} \log(v_i)}{u_i^{-\theta} + v_i^{-\theta} - 1} \right. \\ &\quad \left. - \frac{1}{\theta^2} \sum_{i=1}^N \log(u_i^{-\theta} + v_i^{-\theta} - 1) \right] - \sum_{i=1}^N \log(u_i v_i) \end{aligned}$$

y solo resta resolver  $S(\theta) = 0$  para  $\theta$ .

**Cópula Farlie-Gumbel-Morgenstern** Del método de máxima verosimilitud se tiene que la función de verosimilitud  $L(\theta)$  es,

$$\begin{aligned} L(\theta) &\propto \prod_{i=1}^N c(u_i, v_i; \theta) \\ &\propto \prod_{i=1}^N \left( 1 + \theta(1 - 2u_i)(1 - 2v_i) \right) \end{aligned}$$

entonces, la función log-verosimilitud y la función de Score quedan de la siguiente forma:

$$\begin{aligned} l(\theta) &= \sum_{i=1}^N \log \left( 1 + \theta(1 - 2u_i)(1 - 2v_i) \right) \\ S(\theta) &= \sum_{i=1}^N \frac{(1 - 2u_i)(1 - 2v_i)}{1 + \theta(1 - 2u_i)(1 - 2v_i)} \end{aligned}$$

y resolver para  $\theta$  la expresión  $S(\theta) = 0$

**Cópula de Frank** Del método de máxima verosimilitud se tiene que la función de verosimilitud  $L(\theta)$  es,

$$\begin{aligned} L(\theta) &\propto \prod_{i=1}^N c(u_i, v_i; \theta) \\ &\propto \prod_{i=1}^N \left( \frac{-\theta(e^{-\theta} - 1)e^{-\theta(u_i+v_i)}}{((e^{-\theta} - 1) + (e^{-\theta u_i} - 1)(e^{-\theta v_i} - 1))^2} \right) \end{aligned}$$

y la función log-verosimilitud es:

$$\begin{aligned} l(\theta) &= \sum_{i=1}^N \log \left( \frac{-\theta(e^{-\theta} - 1)e^{-\theta(u_i+v_i)}}{((e^{-\theta} - 1) + (e^{-\theta u_i} - 1)(e^{-\theta v_i} - 1))^2} \right) \\ &= N \log(-\theta) + N \log(e^{-\theta} - 1) - \theta \sum_{i=1}^N (u_i + v_i) \\ &\quad - 2 \sum_{i=1}^N \log(e^{-\theta} - 1 + (e^{-\theta u_i} - 1)(e^{-\theta v_i} - 1)) \end{aligned}$$



y su función de Score es:

$$S(\theta) = -\frac{N}{\theta} - \frac{Ne^{-\theta}}{e^{-\theta} - 1} - \sum_{i=1}^N (u_i + v_i) - 2 \sum_{i=1}^N \frac{-e^{-\theta} - v_i e^{-\theta v_i} (e^{-\theta u_i} - 1) - u_i e^{-\theta u_i} (e^{-\theta v_i})}{(e^{-\theta} - 1 + (e^{-\theta u_i} - 1)(e^{-\theta v_i} - 1))}$$

**Cópula Gaussiana** Del método de máxima verosimilitud se tiene que la función de verosimilitud  $L(\theta)$  es,

$$\begin{aligned} L(\theta) &\propto \prod_{i=1}^N c(u_i, v_i; \theta) \\ &\propto \prod_{i=1}^N \left( \frac{1}{\sqrt{1-\theta^2}} \exp\left\{ -\frac{x_i^2 + y_i^2 - 2\theta x_i y_i}{2(1-\theta^2)} \right\} \right. \\ &\quad \left. \exp\left\{ -\frac{1}{2}(x_i^2 + y_i^2) \right\} \right) \\ &\propto \left( \frac{1}{\sqrt{1-\theta^2}} \right)^N \exp\left\{ \frac{-1}{2(1-\theta^2)} \sum_{i=1}^N (x_i^2 + y_i^2 - 2\theta x_i y_i) \right\} \end{aligned}$$

donde,  $x_i = \Phi^{-1}(u_i)$  y  $y_i = \Phi^{-1}(v_i)$  y la función log-verosimilitud es:

$$\begin{aligned} l(\theta) &= \log \left( \frac{1}{\sqrt{1-\theta^2}} \right)^N \exp\left\{ \frac{-1}{2(1-\theta^2)} \sum_{i=1}^N (x_i^2 + y_i^2 - 2\theta x_i y_i) \right\} \\ &= -\frac{N}{2} \log(1-\theta^2) - \frac{1}{2(1-\theta^2)} \sum_{i=1}^N (x_i^2 + y_i^2 - 2x_i y_i) \end{aligned}$$

y la función de Score que se debe resolver esta dada por:

$$S(\theta) = \frac{2N\theta - 1}{1-\theta^2} - \frac{2\theta}{(1-\theta^2)^2} \sum_{i=1}^N (x_i^2 + y_i^2 - 2x_i y_i) - 2 \sum_{i=1}^N x_i y_i$$

Se han dado las expresiones de la función de Score para cada cópula, el siguiente paso en el método de máxima verosimilitud es igualar a cero la función  $S(\theta)$  y resolver para  $\theta$ , de forma analítica es prácticamente imposible resolver la ecuación, por lo que se recurre a métodos numéricos para estimar el parámetro  $\theta$ . Hay dos formas de hallar el parámetro:

- \* Hallar directamente el valor de  $\theta$  que maximiza la log-verosimilitud  $l(\theta)$ . En la Tabla 5.1 se presentan las expresiones de la función log-verosimilitud que se debe maximizar para cada una de las cópulas.
- \* Hallar la raíz de la función de Score. En la Tabla 5.2 se presenta un resumen de las expresiones de la función de Score para cada cópula.

Con cualquiera de las dos opciones se encuentra el estimador de máxima verosimilitud para  $\theta$ , en la implementación que se hizo se eligió maximizar la función  $l(\theta)$ .

Hemos resuelto uno de los dos problemas planteados (estimación del parámetro de dependencia), ahora veamos como nos enfrentamos al problema de la selección de cópula.

Cópula	Log-Verosimilitud
AMH	$l(\theta) = \sum_{i=1}^N \left( \log [1 + \theta(u_i + v_i + u_i v_i - 2) - \theta^2(u_i + v_i - u_i v_i - 1)] - 3 \log [1 - \theta(1 - u_i)(1 - v_i)] \right)$
Clayton	$l(\theta) = N \log(1 + \theta) - (\theta + 1) \sum_{i=1}^N \log(u_i v_i) - \left( \frac{1}{\theta} + 2 \right) \sum_{i=1}^N \log(u_i^{-\theta} + v_i^{-\theta} - 1)$
FGM	$l(\theta) = \sum_{i=1}^N \log \left( 1 + \theta(1 - 2u_i)(1 - 2v_i) \right)$
Frank	$l(\theta) = N \log(-\theta) + N \log(e^{-\theta} - 1) - \theta \sum_{i=1}^N (u_i + v_i) - 2 \sum_{i=1}^N \log(e^{-\theta} - 1 + (e^{-\theta u_i} - 1)(e^{-\theta v_i} - 1))$
Gaussiana	$l(\theta) = -\frac{N}{2} \log(1 - \theta^2) - \frac{1}{2(1 - \theta^2)} \sum_{i=1}^N (x_i^2 + y_i^2 - 2x_i y_i)$

Tabla 5.1: Función de Log-Verosimilitud para cópulas.

Cópula	Función de Score
AMH	$S(\theta) = \sum_{i=1}^N \left( \frac{u_i + v_i + u_i v_i - 2 - 2\theta(u_i + v_i - u_i v_i - 1)}{1 + \theta(u_i + v_i + u_i v_i - 2) - \theta^2(u_i + v_i - u_i v_i - 1)} \right)$
Clayton	$S(\theta) = \frac{N}{1+\theta} - \left[ \left( \frac{1}{\theta} + 2 \right) \sum_{i=1}^N \frac{-u_i^{-\theta} \log(u_i) - v_i^{-\theta} \log(v_i)}{u_i^{-\theta} + v_i^{-\theta} - 1} - \frac{1}{\theta^2} \sum_{i=1}^N \log(u_i^{-\theta} + v_i^{-\theta} - 1) \right] - \sum_{i=1}^N \log(u_i v_i)$
FGM	$S(\theta) = \sum_{i=1}^N \frac{(1-2u_i)(1-2v_i)}{1+\theta(1-2u_i)(1-2v_i)}$
Frank	$S(\theta) = -\frac{N}{\theta} - \frac{N e^{-\theta}}{e^{-\theta} - 1} - \sum_{i=1}^N (u_i + v_i) - 2 \sum_{i=1}^N \frac{-e^{-\theta} - v_i e^{-\theta} v_i (e^{-\theta} u_i - 1) - u_i e^{-\theta} u_i (e^{-\theta} v_i)}{(e^{-\theta} - 1 + (e^{-\theta} u_i - 1)(e^{-\theta} v_i - 1))}$

Tabla 5.2: Función de Score para cópulas.

## 5.5. Selección de cópula

En el capítulo 4 presentamos algunas de las cópulas que hay en la literatura (las más conocidas y utilizadas ya que han sido las más estudiadas), pero en realidad existen muchas más familias de cópulas, y es aquí donde se presenta otro de los problemas, ¿cómo elegir la mejor cópula?, y es necesario aclarar bajo qué criterio es la mejor de las cópulas o la que más conviene a nuestros intereses. Para ello es necesario definir la cópula empírica, que es una cópula no paramétrica y su construcción se basa en la muestra con la que se trabaja.

### 5.5.1. Cópula empírica

Recordemos un concepto de estadística que nos servirá para la construcción de la cópula empírica.

Suponga que se tiene una muestra de tamaño  $n$ ,  $(X_1, X_2, \dots, X_n)$ . En muchas aplicaciones se dispone de  $n$  datos para los que la suposición de que provienen de una secuencia i.i.d (independiente e idénticamente distribuida) tiene sentido. A menudo interesa obtener el **estadístico de orden**, que no es más que ordenar los datos de menor a mayor, la notación que utilizaremos para tal estadístico es la siguiente:  $x_{(1)}, x_{(2)}, \dots, x_{(n)}$ , donde  $x_{(1)} \leq x_{(2)} \leq \dots \leq$

$x_{(n)}$ .

Ahora presentamos la cópula empírica introducida por Deheuvels (1979, 1981) [8].

**Definición (Cópula Empírica):** Sea  $\{(x_k, y_k)\}_{k=1}^n$  una muestra de tamaño  $n$  de una distribución continua bivariada. La Cópula empírica es la función  $C_n$  dada por:

$$C_n\left(\frac{i}{n}, \frac{j}{n}\right) = \frac{\#\{(x, y) \text{ en la muestra tal que } x \leq x_{(i)} \text{ y } y \leq y_{(j)}\}}{n}$$

donde  $x_{(i)}$  y  $y_{(j)}$ ,  $1 \leq i, j \leq n$  denota el estadístico de orden de la muestra.

Al igual que las funciones de cópula presentadas en el Capítulo 4, la cópula empírica permite la idea de la función de densidad de cópula, llamada cópula empírica de frecuencias.

La cópula empírica de frecuencias es la forma equivalente de la densidad de la cópula, notemos que la cópula empírica no es una función continua por lo que su densidad sólo es una estimación.

La forma de estimar la cópula empírica de frecuencias es:

$$c_n\left(\frac{i}{n}, \frac{j}{n}\right) = \begin{cases} \frac{1}{n}, & \text{si } (x_{(i)}, y_{(j)}) \text{ es un elemento de la muestra,} \\ 0, & \text{en otro caso.} \end{cases}$$

Notemos que la cópula empírica solo depende de la muestra, por lo que el número de observaciones es importante para tener una buena estimación de la cópula empírica, pero un tamaño grande en la muestra puede resultar costoso computacionalmente, por ello es importante hacer una buena elección para el tamaño de la muestra. El algoritmo 3 es para la construcción de la cópula empírica en dimensión dos, y básicamente consiste en los siguientes pasos:

1. Dada la matriz  $A$  de tamaño  $n \times 2$ , que es la muestra en el espacio  $(x, y)$ , comenzamos ordenando por separado cada una de las columnas, y con la función *unique* creamos un nuevo vector para cada variable en el que no se tengan valores repetidos,  $(x_2, y_2)$ , y obtenemos su longitud  $(n_x, n_y)$ , que no necesariamente serán iguales.
2. El siguiente paso es ordenar (de forma ascendente) por renglones la muestra, con base en la primera columna.

3. Posteriormente, contamos el número de parejas ( $c$ ) en la muestra que sean menores a cada par  $(x_2(i), y_2(j))$ , para  $i = 1, \dots, n_x$  y  $j = 1, \dots, n_y$ , se divide entre el tamaño de la muestra y lo reportamos como una entrada de la cópula empírica  $\tilde{C}(i, j) = c/n$ .

---

**Algoritmo 3** Cópula empírica
 

---

Entrada:  $A=(x,y)$  : Muestra en el espacio  $(x,y)$ .  $n$ : Tamaño de la muestra.

```

1:  $x \leftarrow \text{sort}(A(:,1))$ .
2:  $y \leftarrow \text{sort}(A(:,2))$ .
3:  $x_2 \leftarrow \text{unique}(x)$ .
4:  $y_2 \leftarrow \text{unique}(y)$ 
5:  $n_x \leftarrow \text{size}(x_2)$ 
6:  $n_y \leftarrow \text{size}(y_2)$ 
7:  $\text{datos}_{orden} = \text{sortrows}(A)$ ;
8: para  $i = 1, \dots, n_x$  hacer
9:    $x_i \leftarrow x(i)$ 
10:  para  $j = 1, \dots, n_y$  hacer
11:     $k \leftarrow 1$ 
12:     $B = 1$ 
13:    mientras  $B = 1$  hacer
14:       $c \leftarrow 0$ 
15:       $y_j \leftarrow y(j)$ 
16:      si  $\text{datos}_{orden}(k, 1) \leq x_i$  entonces
17:        si  $\text{datos}_{orden}(k, 2) \leq y_j$  entonces
18:           $c \leftarrow c + 1$ 
19:           $k \leftarrow k + 1$ 
20:        si no
21:           $k \leftarrow k + 1$ 
22:        fin si
23:      si no
24:         $B \leftarrow 0$ 
25:      fin si
26:    fin mientras
27:     $\tilde{C}(i, j) \leftarrow \frac{c}{n}$ 
28:  fin para
29: fin para

```

---

Una vez que sabemos como construir la cópula empírica podemos introducir el método de selección de cópulas propuesto por Ané y Kharoubi (2003). Consiste en comparar los valores de la cópula paramétrica con otra no paramétrica. Se considera el caso de la cópula empírica (como la cópula no paramétrica)  $C_e(F_1(Y_1), F_2(Y_2))$  y alguna cópula paramétrica  $\tilde{C}_p(F(Y_1), F_2(Y_2))$  para seleccionar la que mejor se ajusta a los datos, se calcula un estimador de la distancia definido como:

$$Distancia = \sum_{i=1}^n (\hat{C}_{ei} - \tilde{C}_{pi})^2 \quad (5.5)$$

Entonces, dada una muestra y un conjunto de cópulas del que se desea seleccionar una de ellas, lo que se debe hacer es lo siguiente:

1. Estimar la cópula empírica.
2. Proponer una cópula paramétrica.
3. Evaluar la muestra tanto en la empírica como en la paramétrica y calcular la distancia con la expresión 5.5.

Los pasos 2 y 3 se repiten para cada cópula propuesta y finalmente se elige la que tenga la menor distancia. En el algoritmo 4 se presenta el pseudocódigo para la selección de cópulas.

---

**Algoritmo 4** Pseudocódigo para la selección de cópula

---

$n_c$  : Número de cópulas.  $\theta$  : Parámetro de dependencia

- 1:  $\tilde{C} \leftarrow$  cópulas propuestas.
  - 2:  $\hat{C}_e \leftarrow$  Copula\_empirica(X).
  - 3: **para**  $k = 1, \dots, n_c$  **hacer**
  - 4:  $Distancia(k) = \sum_{i=1}^n (\hat{C}_e(X_i) - \tilde{C}_j(X_i; \theta_k))^2$
  - 5: **fin para**
  - 6:  $[In, valor] \leftarrow \min(Distancia_k)$ .
  - 7: Mejor\_copula  $\leftarrow \tilde{C}_{In}$
- 

La parte complicada en la selección de cópulas es la evaluación de la muestra en la cópula empírica, ya que no se tiene una función continua lo que puede generar problemas numéricos.

En éste capítulo se han planteado soluciones para los problemas de la estimación del parámetro y la selección de cópula, pero falta otro paso importante en los EDAs, la etapa de simulación y eso se presenta en el siguiente capítulo.

# Capítulo 6

## Simulación

La generación de nuevos individuos a partir del modelo seleccionado es una etapa importante en los EDAs, ya que se desea tener una muestra que describa correctamente el comportamiento de los mejores individuos de la generación anterior manteniendo la dependencia entre variables. Por ello es importante tener un buen método de simulación. En éste capítulo presentamos los dos métodos utilizados para simular de una cópula [30].

### 6.1. Muestreo condicional

Para muchas cópulas, el muestreo condicional es el método más simple de simulación, consiste en sólo tres pasos a seguir, a continuación se presenta el algoritmo de este método:

---

**Algoritmo 5** Muestreo condicional

---

- 1:  $v_1, v_2 \leftarrow rand()$ .
  - 2:  $u_1 \leftarrow v_1$ .
  - 3:  $v_2 \leftarrow C_2(u_2|u_1 = v_1) = \frac{\partial C(u_1, u_2)}{\partial u_1}$ .
  - 4: Plot  $(u_1, u_2)$
- 

La pareja  $(u_1, u_2)$  son variables distribuidas uniformemente generadas de su correspondiente cópula  $C(u_1, u_2; \theta)$ . Notemos que para el muestreo condicional es necesario tener la función de distribución condicional, que no es más que la derivada parcial de la cópula. Sabemos que una cópula es continua en todo su dominio, así que es posible calcular la primera derivada parcial.



Para que no quede duda sobre el procedimiento del muestreo condicional se presentan algunos ejemplos con las cópulas antes mencionadas.

### 6.1.1. Ejemplos

Aunque hemos presentado el algoritmo para simular de una cópula mediante el muestreo condicional, en muchas ocasiones resulta confuso el procedimiento a seguir, es por ello que presentamos el algoritmo para cada una de las cópulas vistas en el Capítulo 4 a excepción de la cópula Gaussiana.

**Cópula Ali-Mikhail-Haq (AMH).** Recordemos que su función de distribución esta dada por:

$$C(u, v) = uv(1 - \theta(1 - u)(1 - v))^{-1} \quad (6.1)$$

ahora, hallamos su función de distribución condicional  $C_v(v|u)$  derivando (6.1) respecto a  $u$

$$\begin{aligned} C_v(v|u) &= \frac{\partial C(u, v; \theta)}{\partial u} \\ &= \frac{\partial [uv(1 - \theta(1 - u)(1 - v))^{-1}]}{\partial u} \\ &= \frac{-uv\theta(1 - v)}{(1 - \theta(1 - u)(1 - v))^2} + \frac{v}{1 - \theta(1 - u)(1 - v)} \\ &= \frac{-uv\theta(1 - v) + v(1 - \theta(1 - u)(1 - v))}{(1 - \theta(1 - u)(1 - v))^2} \\ &= \frac{v(1 - \theta(1 - v))}{(1 - \theta(1 - u)(1 - v))^2} \end{aligned}$$

Cuando ya se tiene la función de distribución condicional, entonces usamos el algoritmo 6 para generar variables aleatorias de una cópula AMH.

---

**Algoritmo 6** Simulación de datos de una cópula Ali-Mikhail-Haq

---

**Entrada.** Parámetro de dependencia,  $\theta$ .

- 1:  $w_1, w_2 \leftarrow rand()$
  - 2:  $u \leftarrow w_1$ .
  - 3:  $v \leftarrow \frac{2w_2(a\theta-1)^2}{b+\sqrt{c}}$
- donde,

$$a = 1 - u$$

$$b = -\theta(2aw_2 + 1) + 2\theta^2a^2w_2 + 1$$

$$c = \theta^2(4a^2w_2 - 4aw_2 + 1).$$

- 4: Plot  $(u, v)$
- 

**Cópula Clayton** Recordemos la función de distribución de la cópula Clayton:

$$C(u, v; \theta) = \max \left\{ (u^{-\theta} + v^{-\theta} - 1)^{-1/\theta}, 0 \right\}, \quad (6.2)$$

Consideremos el caso en que  $C(u, v; \theta) = (u^{-\theta} + v^{-\theta} - 1)^{-1/\theta}$ , hallamos la función de distribución condicional  $C_v(v|u)$  derivando respecto a  $u$

$$\begin{aligned} C_v(v|u) &= \frac{\partial C(u, v; \theta)}{\partial u} \\ &= \frac{\partial [(u^{-\theta} + v^{-\theta} - 1)^{-1/\theta}]}{\partial u} \\ &= -\frac{1}{\theta} (u^{-\theta} + v^{-\theta} - 1)^{-1-1/\theta} (-\theta)(u^{-\theta-1}) \\ &= u^{-(\theta+1)} (u^{-\theta} + v^{-\theta} - 1)^{-(1+1/\theta)} \end{aligned}$$

cuando ya se tiene la distribución condicional, entonces el algoritmo 7 sirve para simular de una cópula de Clayton y queda de la siguiente manera:

---

**Algoritmo 7** Simulación de datos de una cópula Clayton

---

**Entrada.** Parámetro de dependencia,  $\theta$ .

- 1:  $w_1, w_2 \leftarrow rand()$
  - 2:  $u \leftarrow w_1$ .
  - 3:  $v \leftarrow \left( \frac{1}{u^\theta} (w_2^{-\theta/(\theta+1)} - 1) + 1 \right)^{-1/\theta}$
  - 4: Plot  $(u, v)$
- 

**Cópula Farli-Gumbel-Morgenstern.** Su función de distribución esta dada como:

$$C(u, v) = uv(1 + \theta(1 - u)(1 - v)). \quad (6.3)$$

Al igual que en los casos anteriores, hallamos su función de distribución condicional.

$$\begin{aligned} C_v(v|u) &= \frac{\partial C(u, v; \theta)}{\partial u} \\ &= \frac{\partial [uv(1 + \theta(1 - u)(1 - v))]}{\partial u} \\ &= -uv\theta(1 - v) + v(1 + \theta(1 - u)(1 - v)) \\ &= v + v\theta - v^2\theta - 2uv\theta(1 - v) \\ &= 2uv^2\theta - 2uv\theta - v^2\theta + v\theta + v \\ &= -v^2(\theta(1 - 2u)) + v(1 + \theta(1 - 2u)) \end{aligned}$$

Una vez que tenemos la función de distribución condicional, presentamos el algoritmo 8 para generar variables aleatorias de una cópula FGM:

---

**Algoritmo 8** Simulación de datos de una cópula Farli-Gumbel-Morgenstern

---

**Entrada.** Parámetro de dependencia,  $\theta$ .

- 1:  $w_1, w_2 \leftarrow rand()$
- 2:  $u \leftarrow w_1$ .
- 3:  $v \leftarrow \frac{2w_2}{a+b}$  donde,

$$a = 1 + \theta(1 - 2u)$$

$$b = \sqrt{a^2 - 4(a - 1)w}$$

- 4: Plot  $(u, v)$
-

**Cópula Frank.** Recordemos su función de distribución que esta dada por:

$$C(u, v; \theta) = -\frac{1}{\theta} \log \left( 1 + \frac{(e^{-\theta u} - 1)(e^{-\theta v} - 1)}{e^{-\theta} - 1} \right) \quad (6.4)$$

Derivamos (6.4) respecto a la variable  $u$  para hallar su función de distribución condicional,

$$\begin{aligned} C_v(v|u) &= \frac{\partial C(u, v; \theta)}{\partial u} \\ &= \frac{\partial \left[ -\frac{1}{\theta} \log \left( 1 + \frac{(e^{-\theta u} - 1)(e^{-\theta v} - 1)}{e^{-\theta} - 1} \right) \right]}{\partial u} \\ &= -\frac{1}{\theta} \left( 1 + \frac{(e^{-\theta u} - 1)(e^{-\theta v} - 1)}{e^{-\theta} - 1} \right)^{-1} \left( \frac{-\theta e^{-\theta u} (e^{-\theta v} - 1)}{e^{-\theta} - 1} \right) \end{aligned}$$

En el Algoritmo 9 se muestra el pseudocódigo para simular de una cópula Frank.

---

**Algoritmo 9** Simulación de datos de una cópula Frank

---

**Entrada.** Parámetro de dependencia,  $\theta$ .

- 1:  $w_1, w_2 \leftarrow rand()$
  - 2:  $u \leftarrow w_1$ .
  - 3:  $v \leftarrow -\frac{1}{\theta} \left\{ \log \left[ (1 - w_2)e^{-\theta u} + w_2 e^{-\theta} \right] - \log \left[ (1 - w_2)e^{-\theta u} + w_2 \right] \right\}$
  - 4: Plot  $(u, v)$ .
- 

**Cópula Gumbel.** Su función de distribución esta dada por:

$$C(u, v; \theta) = \exp \left( -(\tilde{u}^\theta + \tilde{v}^\theta)^{1/\theta} \right), \quad (6.5)$$

donde,  $\tilde{u} = -\log(u)$  y  $\tilde{v} = -\log(v)$ . Derivamos (6.5) respecto a  $u$  para obtener su función de distribución condicional y de ahí poder realizar la simulación. Notemos que,

$$\frac{\partial \tilde{u}}{\partial u} = -\frac{1}{u} \quad y \quad \frac{\partial \tilde{v}}{\partial u} = 0$$

$$\begin{aligned}
C_v(v|u) &= \frac{\partial C(u, v; \theta)}{\partial u} \\
&= \frac{\partial \left[ \exp\left(-(\tilde{u}^\theta + \tilde{v}^\theta)^{1/\theta}\right) \right]}{\partial u} \\
&= \exp\left(-(\tilde{u}^\theta + \tilde{v}^\theta)^{1/\theta}\right) \frac{\partial \left[ -(\tilde{u}^\theta + \tilde{v}^\theta)^{1/\theta} \right]}{\partial u} \\
&= C(u, v; \theta) \left(-\frac{1}{\theta}\right) (\tilde{u}^\theta + \tilde{v}^\theta)^{1/\theta-1} (\theta \tilde{u}^{\theta-1}) \left(-\frac{1}{u}\right) \\
&= \frac{C(u, v; \theta)}{u} (\tilde{u}^{\theta-1}) \left(\frac{1}{(\tilde{u}^\theta + \tilde{v}^\theta)^{1/\theta}}\right)^{\theta-1} \\
&= \frac{C(u, v)}{u} \left(\frac{\log(u)}{\log(C(u, v))}\right)^{\theta-1}
\end{aligned}$$

En el Algoritmo 10 se tiene el pseudocódigo para simular de una cópula Gumbel. La diferencia de ésta cópula es que no es posible resolver la ecuación de manera analítica, por lo que se debe resolver numéricamente lo que implica un mayor costo computacional.

---

**Algoritmo 10** Simulación de datos de una cópula Gumbel

---

**Entrada.** Parámetro de dependencia,  $\theta$ .

- 1:  $w_1, w_2 \leftarrow \text{rand}()$
  - 2:  $u \leftarrow w_1$ .
  - 3: Solve  $\left[ w_2 = \frac{C(u, v)}{u} \left(\frac{\log(u)}{\log(C(u, v))}\right)^{\theta-1} \right]$
  - 4: Plot  $(u, v)$ .
- 

En la Tabla 6.1.1 se presenta un resumen de las derivadas parciales de algunas de las cópulas presentadas en el Capítulo 4.

El método condicional es el que hemos utilizado para la simulación de todas las cópulas, excepto para la cópula Gaussiana, ya que no se tiene una expresión cerrada para su función de distribución  $C(u, v)$  y no es posible derivar analíticamente. Por ello recurrimos a otro método de simulación llamado muestreo elíptico.

Cópula	Distribución condicional		Observaciones
AMH	$C_v(v u) = \frac{v(1-\theta(1-v))}{(1-\theta(1-u)(1-v))^2}$	$v = \frac{2w_2(a\theta-1)^2}{b+\sqrt{c}}$	$a = 1 - u;$ $b = -\theta(2aw_2 + 1) + 2\theta^2a^2w_2 + 1;$ $c = \theta^2(4a^2w_2 - 4aw_2 + 1) - \theta(4aw_2 - 4w_2 + 2) + 1;$
Clayton	$C_v(v u) = u^{-(\theta+1)}(u^{-\theta} + v^{-\theta} - 1)^{-(1+1/\theta)}$	$v_2 = C_2(u_2 u_1 = v_1) = \frac{\partial C(u_1, u_2)}{\partial u_1}$	
FGM	$C_v(v u) = -v^2(\theta(1 - 2u)) + v(1 + \theta(1 - 2u))$	$v = \frac{2w_2}{a+b}$	$a = 1 + \theta(1 - 2w);$ $b = \sqrt{a^2 - 4(a - 1)w};$
Frank	$C_v(v u) = \frac{e^{-\theta u}(e^{-\theta v} - 1)}{(e^{-\theta} - 1) + (e^{-\theta u} - 1)(e^{-\theta v} - 1)}$	$v = -\frac{1}{\theta} \left\{ \log \left[ (1 - w_2)e^{-\theta u} + w_2e^{-\theta} \right] - \log \left[ (1 - w_2)e^{-\theta u} + w_2 \right] \right\}$	
Gumbel	$C_v(v u) = \frac{C(u, v)}{u} \left( \frac{\log(u)}{\log(C(u, v))} \right)^{\theta-1}$	Resolver numéricamente $C_v(v u) = w_2$	

Tabla 6.1: Muestreo condicional para cópulas.

## 6.2. Muestreo elíptico

Este método es para simular de distribuciones elípticas como lo son la normal bivariada y la distribución t de Student bivariada. Lo que se hace en éste método es inducir de alguna forma la correlación (de Pearson) de los datos de la muestra en la nueva población, ya que para éste método de simulación, el parámetro de dependencia  $\theta$  de la cópula es la correlación de Spearman. Y precisamente es éste método el que utilizaremos para simular de la cópula Gaussiana.

Los pasos para generar variables  $u_1$  y  $u_2$  de una cópula Gaussiana  $C(u_1, u_2; \theta)$  son los siguientes:

---

**Algoritmo 11** Simulación de datos de una cópula Gaussiana

---

**Entrada.** Parámetro de dependencia,  $\theta$ .

- 1:  $v_1, v_2 \leftarrow randn()$
  - 2:  $y_1 \leftarrow v_1$
  - 3:  $y_2 \leftarrow v_1\theta + \sqrt{1 - \theta^2}$ .
  - 4:  $u_i \leftarrow \Phi(y_i)$  para  $i = 1, 2$  donde  $\Phi$  es la función de distribución acumulada de la densidad normal estándar.
  - 5: Plot  $(u_1, u_2)$
- 

Así, la pareja  $(u_1, u_2)$  son variables aleatorias con distribución uniforme, generadas de la Cópula Gaussiana  $C(\theta)$ .

Las siguientes figuras representan simulaciones de las cópulas presentadas para diferentes parámetros de dependencia, en las que podemos observar que de acuerdo a la cópula y al parámetro de dependencia se generan diferentes estructuras en la muestra.

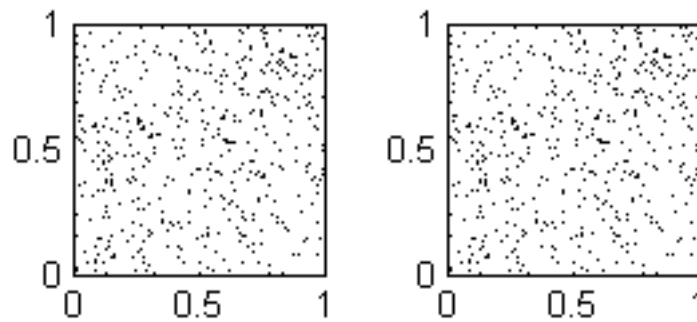


Figura 6.1: Muestreo de la cópula Ali-Mikhail-Haq para  $\theta = 0.5$  (Izquierda) y  $\theta = -0.5$  (Derecha). En ninguna de las muestras se percibe un patrón de dependencia y esto es debido a que ésta cópula no modela correlaciones fuertes.

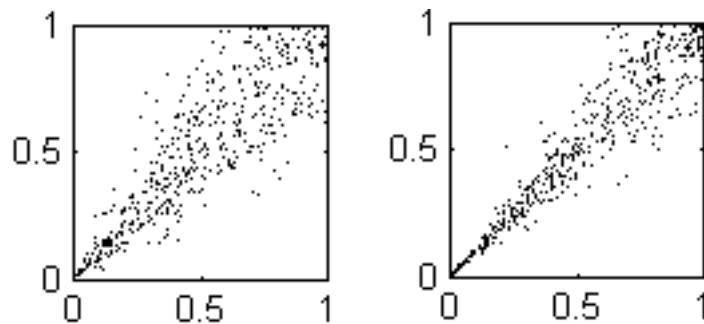


Figura 6.2: Muestreo de la cópula Clayton para  $\theta = 5$  (Izquierda) y  $\theta = 10$  (Derecha). Podemos ver que en ambas muestras se concentra un gran número de puntos en el origen y comienzan a expandirse sobre la diagonal, es por ello que se dice que la cópula Clayton tiene una *cola inferior*.



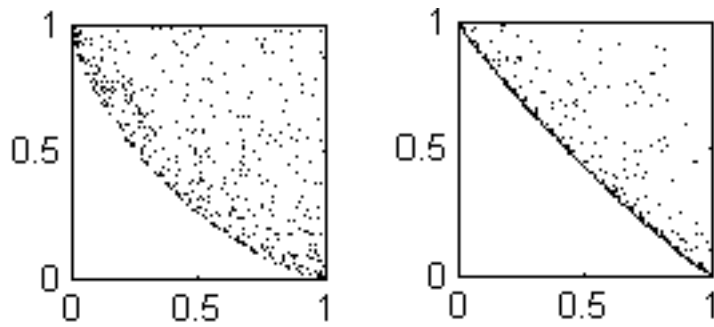


Figura 6.3: Muestreo de la cópula Clayton para  $\theta = -0.7$  (Izquierda) y  $\theta = -0.9$  (Derecha). Notemos que ésta cópula genera un límite, que depende del parámetro de dependencia y debajo de él no genera puntos.

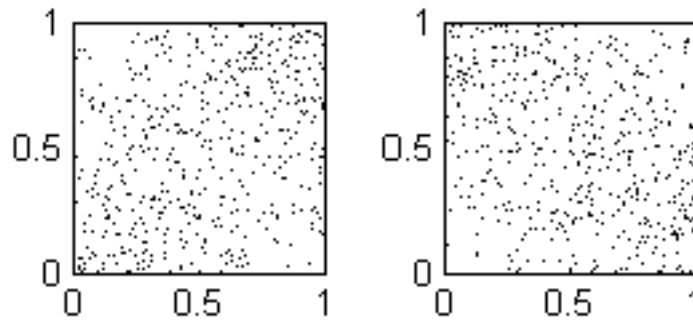


Figura 6.4: Muestreo de la cópula FGM para  $\theta = 0.95$  (Izquierda) y  $\theta = 10$  (Derecha). Ésta cópula modela correlaciones débiles, por lo que las muestra que se generan de ella parecen ser independientes.

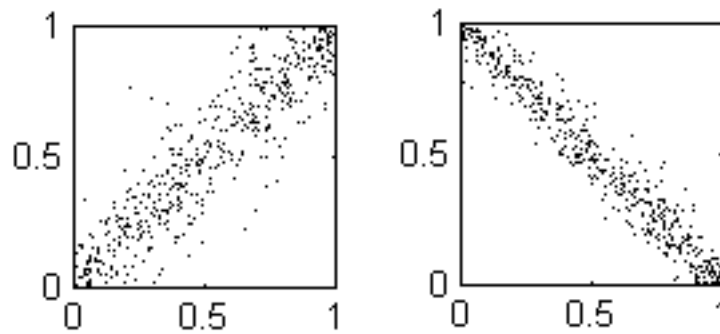


Figura 6.5: Muestreo de la cópula Frank para  $\theta = 15$  (Izquierda) y  $\theta = -20$  (Derecha). Las muestras siempre quedan sobre las diagonales principales del espacio  $[0,1] \times [0,1]$  y dependiendo de su parámetro será la inclinación y el ancho de la banda sobre los que se encuentran los puntos.

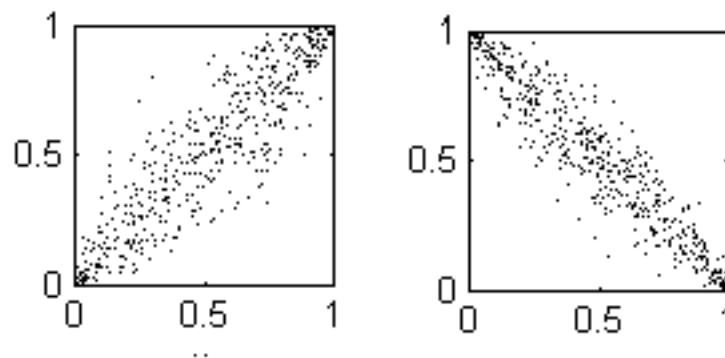


Figura 6.6: Muestreo de la cópula Gaussiana para  $\theta = 0.9$  (Izquierda) y  $\theta = -0.95$  (Derecha). Con ésta cópula los puntos que se simulan forman elipses sobre las diagonales principales cuyo ancho depende del parámetro de la cópula.

Es el momento de hablar sobre las funciones de cópula en los EDAs. En el siguiente capítulo se presentan dos propuestas para trabajar con los EDAs cuyas funciones de distribución son las funciones de cópula.

# Capítulo 7

## EDAs con funciones de cópula

En el Capítulo 3 hablamos sobre los EDAs, algo importante que retomaremos de ese capítulo es el EDA tipo MIMIC. En el Capítulo 4 se habló sobre las funciones de cópula y presentamos algunas de las cópulas más importantes; en este capítulo haremos uso de todo ello y presentaremos dos propuestas para trabajar un problema de optimización utilizando los EDAs. Para ambos casos se construye un EDA tipo MIMIC, la diferencia radica en la función de distribución que se emplea para modelar las interdependencias entre las variables.

La primera propuesta consiste en elegir (sin considerar alguna medida de desempeño) una cópula y trabajar con ella en cada par de variables, en el presente trabajo hemos elegido la cópula Gaussiana, así que construiremos un MIMIC con Cópulas Gaussianas (MIMIC-CG). La segunda propuesta es un MIMIC con selección de cópula (MIMIC-SC) que consiste en considerar algunas cópulas e implementar un método de selección para que en cada par de variables de la cadena MIMIC se elija la cópula que mejor capture la dependencia. Dentro del conjunto de cópulas que proponemos para la selección están la cópula Frank, FGM, Clyton y Gaussiana. Veamos de manera detallada cada una de las propuestas, pero antes haremos el planteamiento general del problema de optimización con el que se trabaja.

### 7.1. Planteamiento

El tipo de problemas de optimización con el que se trabajará son de la siguiente forma,

$$\begin{aligned}
& \text{Minimizar } g(x) && (7.1) \\
& \text{sujeto a} \\
& a_1 \leq x_1 \leq b_1 \\
& a_2 \leq x_2 \leq b_2 \\
& \quad \cdot \\
& \quad \cdot \\
& \quad \cdot \\
& a_d \leq x_d \leq b_d
\end{aligned}$$

donde,

$g(x) : \mathbb{R}^d \rightarrow \mathbb{R}$  es la función objetivo que se debe minimizar.

$x = (x_1, x_2, \dots, x_d)$  es un punto factible en  $\mathbb{R}^d$ .

$a_i$  y  $b_i$ , para  $i = 1, \dots, d$  acotan el espacio de búsqueda para la variable  $x_i$ .

En la sección 7.4 se presentan las funciones de prueba  $g(x)$  con las que se trabajará.

## 7.2. MIMIC con cópulas

Dado un problema de optimización de la forma (7.1), el objetivo es hallar el punto  $x^* = (x_1, x_2, \dots, x_n)$  que minimiza la función objetivo  $g(x)$ . Para ello haremos uso de los Algoritmos de Estimación de Distribución (EDAs), en particular trabajaremos con el MIMIC con representación real, que recordemos es un EDA bivariado en forma de cadena, y para cada par de variables utilizaremos como su función de distribución conjunta una cópula Gaussiana. Hemos elegido trabajar con la cópula Gaussiana ya que modela dependencias grandes, tanto positivas como negativas además de que no es complicada su implementación. En el algoritmo 12 se presenta el pseudocódigo de un MIMIC con cópula Gaussiana (MIMIC-CG), que sigue los siguientes pasos:

1. El primer paso, tanto en el MIMIC como en cualquier EDA, es generar una población inicial  $D_0(X)$  con  $m$  individuos, asumimos independencia entre las variables, así que la población inicial se genera de una distribución uniforme en cada variable.

2. Se obtiene el valor de aptitud de cada individuo evaluando la población inicial en la función de aptitud, en muchas ocasiones es la función que se desea optimizar, es decir, evaluamos  $D_0(X)$  en la función objetivo.
3. Se obtiene una muestra de la población,  $D_0^*(X)$ , seleccionando a los  $n$  ( $n \leq m$ ) mejores individuos en base a su valor de aptitud.
4. Con los individuos de la muestra generar una cadena MIMIC, recordemos rápidamente como generar la cadena, primero hallamos la entropía marginal de cada variable y se elige como nodo raíz a la variable que tenga la menor entropía marginal, y comenzamos la cadena,  $M = \{x^{(1)}\}$ . Posteriormente, se calculan las entropías condicionales  $H(x_j | x^{(1)})$  para  $x_j = \{x_1, x_2, \dots, x_d\} \setminus x^{(1)}$  y se elige como primer hijo a la variable con menor entropía condicional  $M = \{x^{(1)}, x^{(2)}\}$ . La última variable que agregamos a la cadena ahora se toma como el nuevo padre y se repite el procedimiento hasta terminar de enlazar a todas las variables, obteniendo la cadena  $M = \{x^{(1)}, x^{(2)}, \dots, x^{(n)}\}$ .
5. Se obtiene su correlación por rangos  $\tau$  de Kendall y se hace la transformación para hallar el parámetro de dependencia de la cópula, (ver Tabla 4.1).
6. Sabemos que una de las propiedades de las cópulas es que el dominio de las variables es el cuadro unitario  $[0, 1] \times [0, 1]$ , así que debemos hacer el cambio del espacio de las variables  $x_i$  al espacio de las cópulas, y eso lo hacemos mediante la función de distribución acumulada empírica (PDF), entonces el siguiente paso es calcular de manera empírica las funciones de distribución marginal acumulada de las  $d$  variables,  $u_j = F_{x_j}$  para  $j = 1, \dots, d$ , y tenemos la población en el espacio de las cópulas  $D_0^*(U)$ .
7. Posteriormente, una vez que tenemos la cadena MIMIC y las variables  $u_j$ , se debe generar la nueva población y como se mencionó anteriormente, la propuesta es utilizar una cópula Gaussiana para cada par de variables, entonces el procedimiento para simular es el siguiente:
  - a) Tomamos las dos primeras variables de la cadena,  $(x^{(1)}, x^{(2)})$ , y con su parámetro de dependencia  $\theta$ , se simulan  $m$  puntos de la cópula Gaussiana con el parámetro  $\theta$  encontrado, para así obtener los nuevos valores de la población en las columnas correspondientes de

las variables  $(x^{(1)}, x^{(2)})$ , para fines prácticos en la implementación es conveniente tener tanto los valores de la nueva población en términos de las variables  $x$ 's como en términos de  $u$ 's, es decir, guardar  $D'_{k-1}(X)$  y  $D'_{k-1}(U)$ .

- b) Ya que simulamos el primer par de variables de la cadena, el siguiente paso es simular de manera condicional el resto de las variables, entonces se toma el siguiente par de variables de la cadena,  $(x^{(2)}, x^{(3)})$ , pero notemos que en el paso anterior se generaron valores para la variable  $x^{(2)}$ , es por ello que simularemos condicionando sobre los nuevos valores de esta variable, lo que se hace es pasar como argumento la columna de la variable condicionante  $D'_{k-1}(x^{(2)})$ , y al momento de simular con el muestreo elíptico (ver Algoritmo 11) en el paso 1, se asigna esa columna a  $v_1$  y solo se generan los valores para  $v_2$  de una normal estándar y se obtienen los valores de la variable  $x^{(3)}$  en el espacio  $[0,1]$  y se guardan en la población  $D'_{k-1}(u^{(3)})$ , y para pasarla al espacio de las variables  $x$  sera necesario hallar la inversa de la función de distribución acumulada (IPDF) para obtener  $D'_{k-1}(x^{(3)})$ . Se repite hasta haber simulado las  $d$  variables.
8. Una vez que terminamos la simulación, evaluamos nuevamente en la función de aptitud y seleccionamos de  $\{D_k^*(X)\}$  los  $m$  mejores individuos que formarán la siguiente generación  $D_1(X)$ . Se repite el procedimiento hasta alcanzar el criterio de paro que se halla establecido.

### 7.3. MIMIC con selección de cópula

En esta segunda propuesta, la idea es construir nuevamente una cadena MIMIC pero ahora no necesariamente ajustaremos una cópula Gaussiana, sino que propondremos algunas cópulas con las que se desearía trabajar, y posteriormente en base a un criterio seleccionar a la que mejor se ajuste a los datos de la muestra. Para el criterio de selección tomaremos la medida propuesta por Ané y Kharoubi (Ver Capítulo 5). En el Algoritmo 13 se presenta el pseudocódigo de un MIMIC con selección de cópula (MIMIC-SC). A continuación describimos los pasos que se deben seguir para la implementación del MIMIC con selección de cópula, debemos mencionar que el procedimiento

---

**Algoritmo 12** Pseudocódigo de un MIMIC con cópula Gaussiana

---

**Entrada.** Tamaño de la población ( $m$ ), tamaño de la muestra ( $n$ ), dimensión ( $d$ ).

```

1:  $D_0 \leftarrow$  Poblacion_inicial( $m, d$ ).
2:  $k=1$ ;
3: mientras no se cumpla condición de paro hacer
4:    $D_{k-1}^* \leftarrow$  Selecccion ( $D_{k-1}, n$ )
5:    $M \leftarrow$  MIMIC( $D_{k-1}^*$ )
6:    $\tau \leftarrow$  Kendall ( $D_{k-1}^*$ )
7:    $\theta \leftarrow$  sin( $2\pi\tau(M(1), M(2))$ )
8:    $U_{k-1} \leftarrow$  Simulacion_Gaussiana( $M(1), M(2), m, \theta$ )
9:    $D_k^*(:, M(1)) \leftarrow$  IPDF( $U_{k-1}(:, M(1))$ )
10:   $D_k^*(:, M(2)) \leftarrow$  IPDF( $U_{k-1}(:, M(2))$ )
11:  para  $i=2, \dots, d-1$  hacer
12:     $\theta \leftarrow$  sin( $2\pi\tau(M(i), M(i+1))$ )
13:     $U_{k-1} \leftarrow$  Simulacion_CGaussiana( $D_k^*(:, M(i)), M(i+1), m, \theta$ )
14:     $D_k^*(:, M(i+1)) \leftarrow$  IPDF( $U_{k-1}(:, M(i+1))$ )
15:  fin para
16:   $k \leftarrow k + 1$ 
17: fin mientras

```

---



---

**Function** P=Simulacion\_Gaussiana( $i, j, n, \theta$ )

---

```

1: para  $k = 1, \dots, n$  hacer
2:    $v_1, v_2 \leftarrow$  randn()
3:    $y_1 \leftarrow v_1$ 
4:    $y_2 \leftarrow v_1\theta + \sqrt{1 - \theta^2}$ .
5:    $u_l \leftarrow \Phi(y_l)$  para  $l = 1, 2$  donde  $\Phi$  es la función de distribución acumulada de la densidad normal estándar.
6:    $P(k, i) \leftarrow u_1$ 
7:    $P(k, j) \leftarrow u_2$ 
8: fin para

```

---

---

**Function** P=Simulacion\_CGaussiana( $v_1, i, n, \theta$ )

```

1: para j=1,...n hacer
2:    $v_2 \leftarrow randn()$ 
3:    $y_1 \leftarrow v_1(j)$ 
4:    $y_2 \leftarrow y_1\theta + \sqrt{1 - \theta^2}$ .
5:    $u_2 \leftarrow \Phi(y_2)$  donde  $\Phi$  es la función de distribución acumulada de la
   densidad normal estándar.
6:    $P(j, i) \leftarrow u_2$ 
7: fin para

```

---

es similar al MIMIC con cópula Gaussiana, así que la descripción se hará con menos detalle y solo enfatizaremos en los puntos que se crea necesario.

1. Antes que nada, se deben seleccionar  $n_c$  (número de cópulas) cópulas para trabajar, la manera en la que las hemos seleccionado es en base a sus características, nos conviene las cópulas que permiten correlaciones grandes, tanto positivas como negativas, ya que no sabemos el comportamiento que tendrán los datos durante el proceso, además deseamos que la evaluación y la simulación de las cópulas propuestas no sea costoso computacionalmente para la implementación. El número de cópulas que se propone también depende del criterio del investigador.
2. El siguiente paso es generar una población inicial  $D_0(X)$  con  $m$  individuos, asumimos independencia entre las variables, así que la población inicial se genera de una distribución uniforme en cada variable.
3. Se obtiene el valor de aptitud de cada individuo evaluando la población inicial en la función objetivo  $g(x)$ .
4. Se obtiene una muestra de la población,  $D_0^*(X)$ , seleccionando a los  $n$  ( $n \leq m$ ) mejores individuos en base a su valor de aptitud.
5. Con los individuos de la muestra generar una cadena MIMIC,  $M = \{x^{(1)}, x^{(2)}, \dots, x^{(d)}\}$ .
6. Estimar la correlación por rangos  $\tau$  de Kendall y hacer la transformación para obtener el parámetro de dependencia  $\theta$  para cada una de las cópulas que se propuso.



7. Estimar de manera empírica las funciones de distribución marginal acumulada de las  $d$  variables,  $u_j = F_{x_j}$  para  $j = 1, \dots, d$ , para tener la población en el espacio de las cópulas  $D_0^*(U)$ .
8. Posteriormente, una vez que tenemos las cópulas con las que se trabajará, la cadena MIMIC y las variables  $u_j$ , se debe generar la nueva población y como se mencionó anteriormente, la propuesta es elegir de entre las  $n_c$  cópulas propuestas la que mejor se ajuste a los datos de la muestra, para ello retomaremos la medida propuesta por Ané y Kharoubi (2003), que consiste en comparar los valores de una cópula no paramétrica con los valores de una cópula paramétrica, por lo que será necesario construir la cópula empírica para poder hacer la comparación.
9. Para cada pareja consecutiva de variables en la cadena hacer:
  - a) Evaluar la cópula empírica  $\hat{C}_{ei}$  en los puntos de la muestra.
  - b) Con el parámetro  $\theta$  que se obtuvo para cada cópula, evaluar la muestra en cada una de las cópulas propuestas y hallar las distancias

$$d_j = \sum_{i=1}^n (\hat{C}_{ei} - \tilde{C}_{pji})^2$$

donde,

$\hat{C}_{ei}$  es la cópula empírica.

$C_{pji}$  es la evaluación de la muestra en la  $j$ -ésima cópula propuesta.

Se elige la cópula mas cercana a la cópula empírica, es decir, la que tenga el valor mas pequeño en la distancia calculada.

- c) Una vez que se eligió la cópula, se procede a generar nuevos valores para el par de variables con el que se está trabajando, con el algoritmo de simulación que le corresponda,  $D'_{k-1}(X)$ .

Al igual que en el caso de MIMIC con cópula Gaussiana, la primera pareja de variables se simula con el algoritmo correspondiente a la cópula seleccionada sin modificaciones, pero a partir de la segunda pareja se debe simular condicionando a los nuevos valores que se obtuvieron de la variable que ahora es el padre.

Se debe tener cuidado al momento de pasar los valores ya simulados, ya que en el caso de elegir una cópula Gaussiana se deben pasar los

valores en el espacio de las  $x$ 's y para cualquier otra cópula se pasan los valores de las variables pero en el espacio de las cópulas, es decir, los valores en  $u$ 's.

10. Una vez que se han simulado todas las variables, hallar su la inversa de su función de distribución, evaluamos nuevamente en la función de aptitud y seleccionamos de  $\{D_k^*(X)\}$  los  $m$  mejores individuos que formarán la siguiente generación  $D_1(X)$ . Se repite el procedimiento hasta alcanzar el criterio de paro que se halla establecido.

## 7.4. Experimentos y resultados

Comenzaremos esta sección presentando las cinco funciones de prueba (funciones objetivo) que se eligieron para trabajar. El objetivo es minimizar cada una de ellas con las dos propuestas que se hicieron, MIMIC con cópula Gaussiana y MIMIC con selección de cópula, para poder comparar resultados.

### 7.4.1. Funciones de Prueba

Las funciones de prueba que utilizamos para los experimentos son las siguientes:

#### 1.- Sphere Minimizar

$$g(x) = \sum_{i=1}^n x_i^2$$

Rango de búsqueda  $-600 \leq x_i \leq 600$ . Óptimo global  $x_i^* = 0$  y  $f(x_i^*) = 0$

#### 2.- Ackley Minimizar

$$g(x) = -20e^{-0.2\sqrt{\frac{1}{n}\sum_{i=1}^n x_i^2 - e^{\frac{1}{n}\sum_{i=1}^n \cos(2\pi x_i)}}} + 20 + e^1$$

Rango de búsqueda  $-10 \leq x_i \leq 10$ . Óptimo global  $x_i^* = 0$  y  $f(x_i^*) = 0$

---

**Algoritmo 13** Pseudocódigo de un MIMIC con selección de cópula

---

**Entrada.** Tamaño de la población ( $m$ ), tamaño de la muestra ( $n$ ), dimensión ( $d$ ), número de cópulas ( $n_c$ ), Vector de cópulas con las que se trabajará ( $\tilde{C}$ ).

```

1:  $D_0 \leftarrow$  Poblacion_inicial ( $m,d$ ).
2:  $k=1$ ;
3: mientras no se cumpla condición de paro hacer
4:    $D_{k-1}^* \leftarrow$  Seleccion( $D_{k-1},n$ )
5:    $M \leftarrow$  MIMIC( $D_{k-1}^*$ )
6:    $\tau \leftarrow$  kendall( $D_{k-1}^*$ )
7:    $\hat{C}_e \leftarrow$  Crear_Copula_empirica( $D_{k-1}^*, M(1), M(2)$ )
8:   Muestra  $\leftarrow$  [ $D_{k-1}^*(:, M(1)), D_{k-1}^*(:, M(2))$ ]
9:    $\theta_j(:, :) \leftarrow$  Param_Cop( $\tau$ )
10:  Mejor_copula  $\leftarrow$  Seleccion_copula( $\hat{C}_e, \tilde{C}, n_c, Muestra, \theta, M(1), M(2)$ )
11:   $U_{k-1} \leftarrow$  Simulacion_Mejor_copula( $M(1), M(2), m, \tau(M(1), M(2))$ )
12:   $D_k^*(:, M(1)) \leftarrow$  IPDF( $U_{k-1}(:, M(1))$ )
13:   $D_k^*(:, M(2)) \leftarrow$  IPDF( $U_{k-1}(:, M(2))$ )
14:  para  $i=2, \dots, d-1$  hacer
15:     $\hat{C}_{ei} \leftarrow$  Crear_Copula_empirica( $D_{k-1}^*, M(i), M(i+1)$ )
16:    Muestra  $\leftarrow$  [ $D_{k-1}^*(:, M(i)), D_{k-1}^*(:, M(i+1))$ ]
17:    Mejor_copula  $\leftarrow$  Seleccion_copula( $\hat{C}_e, \tilde{C}, n_c, Muestra, \theta, M(i), M(i+1)$ )
18:    si Mejor_copula == Gaussiana entonces
19:       $U_{k-1} \leftarrow$  Simulacion_CGaussiana( $M(i+1), m, \tau(M(i), M(i+1)), D_k^*(:, M(i))$ )
20:    si no
21:       $U_{k-1} \leftarrow$  Simulacion_CMejor_copula( $M(i+1), m, \tau, U_{k-1}(:, M(i))$ )
22:    fin si
23:     $D_k^*(:, M(i+1)) \leftarrow$  IPDF( $U_{k-1}(:, i+1)$ )
24:  fin para
25:   $k \leftarrow k+1$ 
26: fin mientras

```

---

---

**Function** Mejor\_copula=Seleccion\_copula( $\hat{C}_e, \tilde{C}, n_c, X_i, \theta, i, j$ )

- 1: **para**  $k = 1, \dots, n_c$  **hacer**
  - 2:  $Distancia(k) = \sum_{i=1}^n (\hat{C}_e(X_i) - \tilde{C}_k(X_i; \theta_k))^2$
  - 3: **fin para**
  - 4:  $[In, valor] \leftarrow \min(Distancia_k)$ .
  - 5: Mejor\_copula  $\leftarrow \tilde{C}_{In}$
- 

---

**Function** P=Simulacion\_Clayton(i,j,n, $\tau$ )

- 1:  $\theta \leftarrow \frac{2\tau}{1-\tau}$
  - 2: **para**  $k=1, \dots, n$  **hacer**
  - 3:  $w_1, w_2 \leftarrow rand()$
  - 4:  $u \leftarrow w_1$ .
  - 5:  $v \leftarrow \left( \frac{1}{u^\theta} (w_2^{-\theta/(\theta+1)} - 1) + 1 \right)^{-1/\theta}$
  - 6:  $P(k, i) \leftarrow u$
  - 7:  $P(k, j) \leftarrow v$
  - 8: **fin para**
- 

---

**Function** P=Simulacion\_FGM(i,j,n, $\tau$ )

- 1:  $\theta \leftarrow \frac{9}{2}\tau$
  - 2: **para**  $k=1, \dots, n$  **hacer**
  - 3:  $w_1, w_2 \leftarrow rand()$
  - 4:  $u \leftarrow w_1$ .
  - 5:  $v \leftarrow \frac{2w_2}{a+b}$  donde,
 
$$a = 1 + \theta(1 - 2u)$$

$$b = \sqrt{a^2 - 4(a - 1)w_2}$$
  - 6:  $P(k, i) \leftarrow u$
  - 7:  $P(k, j) \leftarrow v$
  - 8: **fin para**
-

---

**Function** P=Simulacion\_Frank(i,j,n, $\tau$ )

1: Solve( $\tau = 1 - \frac{4}{\theta}[1 - D_1(\theta)]$ )  
 2: **para** k=1,...,n **hacer**  
 3:  $w_1, w_2 \leftarrow rand()$   
 4:  $u \leftarrow w_1$ .  
 5:  $v \leftarrow -\frac{1}{\theta} \left\{ \log \left[ (1 - w_2)e^{-\theta u} + w_2e^{-\theta} \right] - \log \left[ (1 - w_2)e^{-\theta u} + w_2 \right] \right\}$   
 6:  $P(k, i) \leftarrow u$   
 7:  $P(k, j) \leftarrow v$   
 8: **fin para**  
 $D_1(\theta) = \frac{1}{\theta} \int_0^\theta \frac{t}{e^t - 1} dt$

---



---

**Function** P=Simulacion\_Gaussiana(i,j,n, $\tau$ )

1:  $\theta \leftarrow \sin(2\pi\tau)$   
 2: **para** k=1,...,n **hacer**  
 3:  $v_1, v_2 \leftarrow randn()$   
 4:  $y_1 \leftarrow v_1$   
 5:  $y_2 \leftarrow v_1\theta + \sqrt{1 - \theta^2}$ .  
 6:  $u_l \leftarrow \Phi(y_l)$  para  $l = 1, 2$  donde  $\Phi$  es la función de distribución acumulada de la densidad normal estándar.  
 7:  $P(k, i) \leftarrow u_1$   
 8:  $P(k, j) \leftarrow u_2$   
 9: **fin para**

---



---

**Function** P=Simulacion\_CClayton(u,j,n, $\tau$ )

1:  $\theta \leftarrow \frac{2\tau}{1-\tau}$   
 2: **para** k=1,...,n **hacer**  
 3:  $w_2 \leftarrow rand()$   
 4:  $v \leftarrow \left( \frac{1}{u^{(k)\theta}} (w_2^{-\theta/(\theta+1)} - 1) + 1 \right)^{-1/\theta}$   
 5:  $P(k, j) \leftarrow v$   
 6: **fin para**

---

---

**Function** P=Simulacion\_CFGM(u,j,n, $\tau$ )

- 1:  $\theta \leftarrow \frac{9}{2}\tau$
  - 2: **para** k=1,...,n **hacer**
  - 3:  $w_2 \leftarrow rand()$
  - 4:  $v \leftarrow \frac{2w_2}{a+b}$  donde,
 
$$a = 1 + \theta(1 - 2u(k))$$

$$b = \sqrt{a^2 - 4(a - 1)w_2}$$
  - 5:  $P(k, j) \leftarrow v$
  - 6: **fin para**
- 

---

**Function** P=Simulacion\_CFrank(u,j,n, $\tau$ )

- 1: Solve( $\tau = 1 - \frac{4}{\theta}[1 - D_1(\theta)]$ )
  - 2: **para** k=1,...,n **hacer**
  - 3:  $w_2 \leftarrow rand()$
  - 4:  $v \leftarrow -\frac{1}{\theta} \left\{ \log \left[ (1 - w_2)e^{-\theta u(k)} + w_2 e^{-\theta} \right] - \log \left[ (1 - w_2)e^{-\theta u(k)} + w_2 \right] \right\}$
  - 5:  $P(k, j) \leftarrow v$
  - 6: **fin para**
- $$D_1(\theta) = \frac{1}{\theta} \int_0^\theta \frac{t}{e^t - 1} dt$$
- 

---

**Function** P=Simulacion\_CGaussiana( $v_1, i, n, \theta$ )

- 1:  $\theta \leftarrow \sin(2\pi\tau)$
  - 2: **para** k=1,...,n **hacer**
  - 3:  $v_2 \leftarrow randn()$
  - 4:  $y_1 \leftarrow v_1(k)$
  - 5:  $y_2 \leftarrow y_1\theta + \sqrt{1 - \theta^2}$ .
  - 6:  $u_2 \leftarrow \Phi(y_2)$  donde  $\Phi$  es la función de distribución acumulada de la densidad normal estándar.
  - 7:  $P(k, i) \leftarrow u_2$
  - 8: **fin para**
-

**3.- Griewangk** Minimizar

$$g(x) = 1 + \frac{1}{4000} \sum_{i=1}^n x_i^2 - \prod_{i=1}^n \cos\left(\frac{x_i}{\sqrt{i}}\right)$$

Rango de búsqueda  $-600 \leq x_i \leq 600$ . Óptimo global  $x_i^* = 0$  y  $f(x_i^*) = 0$

**4.- Rastrigin** Maximizar

$$g(x) = 10 * d + \sum_{i=1}^n x_i^2 - 10 \cos(2\pi x_i).$$

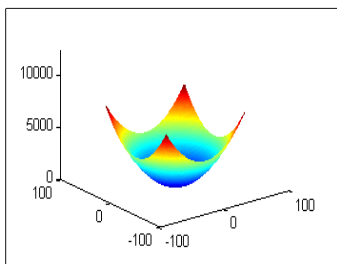
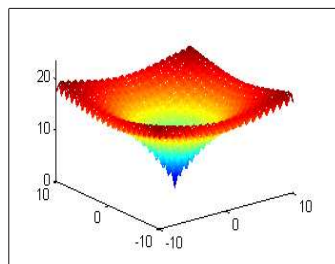
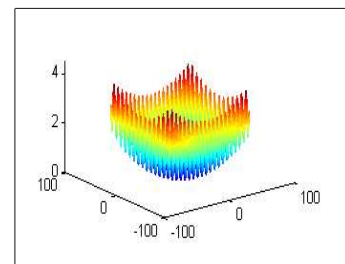
Rango de búsqueda  $-5,12 \leq x_i \leq 5,12$ . Óptimo global  $x_i^* = 0$  y  $f(x_i^*) = 0$

**5.- Rosenbrock** Minimizar

$$g(x) = \sum_{i=1}^{n-1} [100(x_{i+1} - x_i^2)^2 + (x_i - 1)^2].$$

Rango de búsqueda  $-10 \leq x_i \leq 10$ . Óptimo global  $x_i^* = 1$  y  $f(x_i^*) = 0$

En las Figuras 7.1, se tienen las gráficas de las funciones de prueba en dimensión dos.

*Esfera**Ackley**Griewangk*

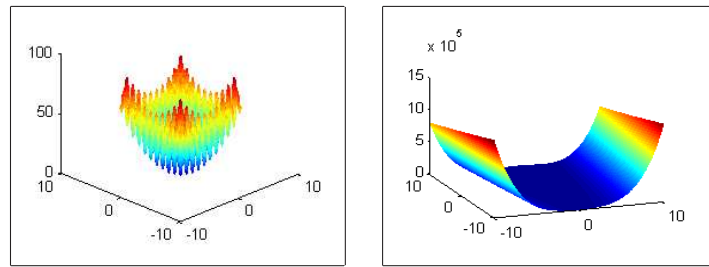
*Rastrigin**Rosenbrock*

Figura 7.1: Funciones de prueba en dimensión dos

### 7.4.2. Implementación y resultados

Es necesario hacer algunas observaciones acerca de la implementación que se hizo del MIMIC-CG y MIMIC-SC:

- . Se aplica selección por truncamiento, que consiste en seleccionar a los mejores individuos en base a su valor de aptitud.
- . El tamaño de la muestra que se selecciona es del 50 % de la población total.
- . Se aplica elitismo del 10 %, es decir, se conserva el 10 % de los mejores individuos en la siguiente generación.
- . Para la implementación del MIMIC-SC, se consideran cuatro cópulas: Gaussiana, Frank, FGM y Clayton. ¿Porqué se eligieron esas cópulas? Una de las razones es que, a excepción de la cópula FGM, permiten medidas de correlación por rangos en  $[-1,1,]$  además de que la evaluación y simulación de estas cópulas no es complicado. Otro aspecto importante son las colas de cada cópula. (con colas nos referimos a las elevaciones que presentan en su función de densidad cerca de los vértices de su dominio, dependiendo del signo de su parámetro de dependencia) Las colas de una cópula permiten generar un mayor número de puntos debajo de ellas, en las Figuras 6.2, 6.3, 6.4, 6.5 y 6.6 se puede observar que el muestreo de cada cópula es diferente aún cuando tengan un mismo valor en su parámetro de dependencia.



- . Los criterios de paro que se utilizan son: Número máximo de evaluaciones de función, cuando en las últimas  $N_m$  generaciones el valor del mejor individuo no presenta cambios respecto al mejor individuo de la generación anterior o cuando su diferencia es pequeña.  $|Mejor(k-1) - Mejor(k)| \leq \varepsilon$
- . El tamaño de la población que se eligió es de 100 individuos.

Los parámetros que se eligieron para las pruebas son los siguientes:

1. Tamaño de la población,  $m = 100$ .
2. Tamaño de la muestra  $n = 50$ .
3. Número máximo de evaluaciones:  $Evaluaciones = 40000$ .
4. Uno de los criterios de paro es cuando la diferencia entre la aptitud del mejor individuo de la generación actual y la anterior sea menor a un valor de tolerancia previamente establecido, para nuestras pruebas tomamos  $tol = 10^{-6}$ .
5. No. de generaciones en las que el mejor individuo no cambia:  $N_m = 30$ .
6. Se realizan 20 corridas de cada experimento para obtener el mejor, el peor, la media, la mediana y la desviación estándar de los resultados que se obtienen.

En la tabla 7.1 se presentan los resultados que se obtiene para cada función de prueba, con el MIMIC-CG, se reporta sobre el valor de la función de aptitud el mejor individuo, su media, la mediana y la desviación estándar.

MIMIC con cópula Gaussiana					
Valor de aptitud					
	Mejor	Peor	Media	Mediana	Desv
<b>Esfera</b>					
D = 2	$6.5000E^{-7}$	$8.4830E^{-5}$	$4.0049E^{-5}$	$3.6985E^{-5}$	$2.8079E^{-5}$
D = 5	$2.9290E^{-5}$	$9.5740E^{-5}$	$5.6025E^{-5}$	$5.4600E^{-5}$	$1.8839E^{-5}$
D = 10	$3.9410E^{-5}$	$9.8720E^{-5}$	$7.2292E^{-5}$	$7.6065E^{-5}$	$1.7252E^{-5}$
<b>Ackley</b>					
D = 2	$9.6660E^{-5}$	0.0010	$4.3739E^{-4}$	$3.6283E^{-4}$	$2.5845E^{-4}$
D = 5	0.0022	0.0040	0.0029	0.0029	$5.6983E^{-4}$
D = 10	0.0037	0.0071	0.0059	0.0059	$9.9342E^{-4}$
<b>Griewangk</b>					
D = 2	$3.3400E^{-6}$	0.0060	0.0010	$4.8110E^{-5}$	0.0018
D = 5	$3.3430E^{-5}$	0.1823	0.0829	0.0971	0.0602
D = 10	$3.3280E^{-5}$	0.4010	0.0205	$7.6770E^{-5}$	0.0896
<b>Rastrigin</b>					
D = 2	$1.6000E^{-6}$	$8.4730E^{-5}$	$3.2193E^{-5}$	$2.1280E^{-5}$	$2.5440E^{-5}$
D = 5	$1.5835E^{-4}$	8.1734	1.1782	$8.0985E^{-4}$	2.4024
D = 10	0.0048	29.2877	8.6986	1.0019	10.9563
<b>Rosenbrock</b>					
D = 2	$6.1700E^{-6}$	$9.7150E^{-5}$	$4.2647E^{-5}$	$3.3805E^{-5}$	$2.9981E^{-5}$
D = 5	$1.2415E^{-4}$	0.0010	$4.7186E^{-4}$	$4.6850E^{-4}$	$2.5151E^{-4}$
D = 10	6.1687	6.7281	6.3887	6.3576	0.1527

Tabla 7.1: Resultados del valor de aptitud con el MIMIC-CG

En la tabla 7.2, se reportan los resultados para el número de evaluaciones de función, nuevamente con el MIMIC-CG.

<b>MIMIC con cópula Gaussiana</b>					
Evaluaciones de función					
	Mejor	Peor	Media	Mediana	Desv
<b>Esfera</b>					
D = 2	1300	8100	1930	1400	1530.1
D = 5	2800	3100	2975	3000	96.6546
D = 10	4500	4900	4740	4750	99.4723
<b>Ackley</b>					
D = 2	3900	10600	6705	6350	2010.4
D = 5	5600	11200	7260	6650	1628.5
D = 10	6800	10700	8350	8300	1430.7
<b>Griewangk</b>					
D = 2	3600	7800	5785	5800	1242.8
D = 5	5500	16800	9010	8100	2956.7
D = 10	4800	12900	6100	5650	1758.0
<b>Rastrigin</b>					
D = 2	2000	4000	2725	2600	555.2382
D = 5	3900	24300	11870	11400	4592.6
D = 10	4000	19700	11765	13250	4335.7
<b>Rosenbrock</b>					
D = 2	1200	3900	2210	1800	795.9767
D = 5	21300	34300	27925	28050	3685.7
D = 10	40100	40100	40100	40100	0

Tabla 7.2: Resultados de el número de evaluaciones de función con el MIMIC-CG

Ahora para el MIMIC-SC, en las tablas 7.3 y 7.4 se presentan los resultados para el valor de aptitud y el número de iteraciones, respectivamente, para las cinco funciones de prueba propuestas.

<b>MIMIC con selección de cópula</b>					
Valor de aptitud					
	Mejor	Peor	Media	Mediana	Desv
<b>Esfera</b>					
D = 2	$1.5300E^{-6}$	$8.6640E^{-5}$	$3.6005E^{-5}$	$3.3940E^{-5}$	$2.5764E^{-5}$
D = 5	$1.6490E^{-5}$	$9.9380E^{-5}$	$6.4961E^{-5}$	$6.2665E^{-5}$	$2.2307E^{-5}$
D = 10	$3.8090E^{-5}$	6.6900	0.3346	$8.7645E^{-5}$	1.4959
<b>Ackley</b>					
D = 2	$8.9100E^{-6}$	$9.7110E^{-5}$	$6.3709E^{-5}$	$7.0710E^{-5}$	$2.5209E^{-5}$
D = 5	$5.5360E^{-5}$	0.0011	$2.4672E^{-4}$	$1.0727E^{-4}$	$2.9134E^{-4}$
D = 10	0.0017	0.0071	0.0046	0.0047	0.0018
<b>Griewangk</b>					
D = 2	$9.4700E^{-6}$	0.0887	0.0074	$3.2391E^{-4}$	0.0195
D = 5	0.0074	0.0740	0.0240	0.0197	0.0177
D = 10	$1.3930E^{-5}$	0.0517	0.0157	0.0135	0.0144
<b>Rastrigin</b>					
D = 2	$9.9600E^{-6}$	$8.7400E^{-5}$	$4.2788E^{-5}$	$4.3715E^{-5}$	$2.4325E^{-5}$
D = 5	$8.1500E^{-6}$	$9.5330E^{-5}$	$5.7293E^{-5}$	$5.6420E^{-5}$	$2.6292E^{-5}$
D = 10	0.0011	2.9859	0.9413	0.9955	0.9391
<b>Rosenbrock</b>					
D = 2	$9.6300E^{-6}$	0.3626	0.0185	$5.3010E^{-5}$	0.0810
D = 5	0.0283	4.3342	1.1132	1.0888	0.8803
D = 10	0.0333	7.3469	5.5187	6.4450	2.2959

Tabla 7.3: Resultados del valor de aptitud con el MIMIC-SC

MIMIC con selección de cópula					
Evaluaciones de función					
	Mejor	Peor	Media	Mediana	Desv
<b>Esfera</b>					
D = 2	1400	8100	2265	1850	1538.4
D = 5	2900	31100	6885	3650	7566.8
D = 10	5200	40100	9395	6600	8349.4
<b>Ackley</b>					
D = 2	2100	7500	3735	3200	1505.5
D = 5	7900	24000	15975	17100	4336.5
D = 10	6800	20700	10950	10050	3845.4
<b>Griewangk</b>					
D = 2	2400	12600	6420	6350	2340.4
D = 5	7900	40100	26070	26050	9642.1
D = 10	5900	26800	14360	14250	6200.0
<b>Rastrigin</b>					
D = 2	1100	5100	2100	1800	917.6629
D = 5	4700	12100	7500	7400	2211.9
D = 10	10000	28700	15760	15250	4619.0
<b>Rosenbrock</b>					
D = 2	1700	16400	6035	4300	4117.9
D = 5	9300	40100	38165	40100	7019.4
D = 10	36100	40100	39830	40100	931.9476

Tabla 7.4: Resultados de el número de evaluaciones de función con el MIMIC-SC

Para facilitar la comparación de los resultados obtenidos con las dos propuestas realizadas se presenta la tabla 7.5, en la que se resume la información. De la columna dos a la siete se tienen los resultados para el mejor, la mediana y la desviación estándar de el valor de aptitud obtenidos con el MIMIC-CG y MIMIC-SC, y en las columnas ocho y nueve se encuentra la mediana de el número de evaluaciones de función para cada una de las propuestas.

MIMIC-CG vs MIMIC-SC								
	Mejor <sub>Apt</sub>		Mediana <sub>Apt</sub>		Desv <sub>Apt</sub>		Mediana <sub>Ev_fun</sub>	
	CG	SC	CG	SC	CG	SC	CG	SC
<b>Esfera</b>								
D = 2	$6.5000E^{-7}$	$1.5300E^{-6}$	$3.6985E^{-5}$	$3.3940E^{-5}$	$2.8079E^{-5}$	$2.5764E^{-5}$	1400	1850
D = 5	$2.9290E^{-5}$	$1.6490E^{-5}$	$5.4600E^{-5}$	$6.2665E^{-5}$	$1.8839E^{-5}$	$2.2307E^{-5}$	3000	3650
D = 10	$3.9410E^{-5}$	$3.8090E^{-5}$	$7.6065E^{-5}$	$8.7645E^{-5}$	$1.7252E^{-5}$	1.4959	6350	6600
<b>Ackley</b>								
D = 2	$9.6660E^{-5}$	$8.9100E^{-6}$	$3.6283E^{-4}$	$7.0710E^{-5}$	$2.5845E^{-4}$	$2.5209E^{-5}$	6350	3200
D = 5	0.0022	$5.5360E^{-5}$	0.0029	$1.0727E^{-4}$	$5.6983E^{-4}$	$2.9134E^{-4}$	6650	17100
D = 10	0.0037	0.0017	0.0059	0.0047	$9.9342E^{-4}$	0.0018	8300	10050
<b>Griewangk</b>								
D = 2	$3.3400E^{-6}$	$9.4700E^{-6}$	$4.8110E^{-5}$	$3.2391E^{-4}$	0.0018	0.0195	5800	6350
D = 5	$3.3430E^{-5}$	0.0074	0.0971	0.0197	0.0602	0.0177	8100	26050
D = 10	$3.3280E^{-5}$	$1.3930E^{-5}$	$7.6770E^{-5}$	0.0135	0.0896	0.0144	5650	14250
<b>Rastrigin</b>								
D = 2	$1.6000E^{-6}$	$9.9600E^{-6}$	$2.1280E^{-5}$	$4.3715E^{-5}$	$2.5440E^{-5}$	$2.4325E^{-5}$	2600	1800
D = 5	$1.5835E^{-4}$	$8.1500E^{-6}$	$8.0985E^{-4}$	$5.6420E^{-5}$	2.4024	$2.6292E^{-5}$	11400	7400
D = 10	0.0048	0.0011	1.0019	0.9955	10.9563	0.9391	13250	15250
<b>Rosenbrock</b>								
D = 2	$6.1700E^{-6}$	$9.6300E^{-6}$	$3.3805E^{-5}$	$5.3010E^{-5}$	$2.9981E^{-5}$	0.0810	1800	4300
D = 5	$1.2415E^{-4}$	0.0283	$4.6850E^{-4}$	1.0888	$2.5151E^{-4}$	0.8803	28050	40100
D = 10	6.1687	0.0333	6.3576	6.4450	0.1527	2.2959	40100	40100

Tabla 7.5: Comparación de resultados MIMIC-CG vs MIMIC-SC.

En la Tabla 7.5 podemos notar que el MIMIC con cópula Gaussiana resulta tener un mejor comportamiento que el MIMIC con Selección de Cópula, aunque en algunos casos, el MIMIC-SC supera los resultados obtenidos con el MIMIC-CG, por ejemplo veamos el caso de la función Ackley en dimensión dos, ambos algoritmos obtienen un valor de aptitud aceptable, sin embargo la mediana de las evaluaciones de función para el MIMIC-SC es menor que para el MIMIC-CG, en esa misma función pero en dimensiones 5 y 10, el MIMIC-SC obtiene mejores valores de aptitud sin embargo lo hace en un mayor número de evaluaciones de función. Otro caso en el que el MIMIC-SC supera al MIMIC-CG es en la función Rastrigin en dimensión 2 y 5, en las que logra bajar la mediana en evaluaciones de función y obtiene buenos resultados para el valor de aptitud. En la función Griewangk, podemos ver que el MIMIC-CG se lleva por completo la prueba, en todos los casos llega a un buen valor de aptitud y disminuye, para dimensión 5 y 10, más del 50% el número de evaluaciones de función.

# Capítulo 8

## Mezclas finitas de Cópulas Gaussianas

Cuando comenzamos a trabajar con funciones de cópula, surgió la idea de trabajar con mezclas de cópulas gaussianas en un EDA, sin embargo a medida que avanzamos encontramos algunos inconvenientes, el más importante fue la inicialización correcta de parámetros. En este capítulo presentamos el avance que logramos en cuanto a mezclas de cópulas, que aunque no se llegó a una implementación en un EDA, logramos estimar parámetros tanto de las cópulas como de las proporciones en la mezcla. Iniciamos el capítulo presentando la definición de una mezcla de funciones.

Un modelo de mezclas finitas consiste en una función de densidad de probabilidad que es una combinación convexa de varias funciones de densidad. Si la variable aleatoria  $Y$  es una mezcla de  $k$  componentes, entonces la función de densidad de probabilidad de  $Y$ ,  $f_Y(y; \Theta)$ , es una suma de densidades con distintos pesos  $\pi_i$  de los componentes. Este modelo se representa de la siguiente manera:

$$f_Y(y; \Theta) = \sum_{i=1}^k \pi_i f_i(y; \theta_i)$$

donde

$\pi_i$  es la proporción del componente  $i$  en la mezcla, y deben cumplir que,

$$0 \leq \pi_i \leq 1; \quad \sum_{i=1}^k \pi_i = 1$$

$f_i(y; \theta_i)$  representa una función de densidad cualquiera con variable  $y$  y un conjunto de parámetros que denotamos con  $\theta_i$ .

Las  $k$  distribuciones no necesariamente son idénticas pero sí deben estar definidas en el mismo espacio al que pertenece el vector aleatorio  $y$ .

Ahora, si consideramos que las funciones de distribución  $f_i$  están definidas en términos de cópulas, en particular consideremos que las  $f_i = c_i(w; \theta_i)$  son cópulas Gaussianas donde  $w$  representa el vector de variables dado como  $w = (u, v)'$ , entonces el modelo que obtenemos para una mezcla finita de cópulas Gaussianas está dado como sigue:

$$c^*(w; \Theta) = \sum_{i=1}^k \pi_i c_i(w; \theta_i) \quad (8.1)$$

donde,

$$w = (u = \Phi(x), v = \Phi(y))$$

$$c_i(w; \theta_i) = \frac{1}{\sqrt{1-\theta_i^2}} \exp\left\{ -\frac{x^2+y^2-2\theta_i xy}{2(1-\theta_i^2)} \right\} \exp\left\{ -\frac{1}{2}(x^2 + y^2) \right\} \text{ es la } i\text{-ésima cópula gaussiana.}$$

Entonces, para realizar inferencia de parámetros en este modelo es conveniente que los datos observados  $w = (u, v)'$  sean un vector aleatorio extraído de cada una de las  $k$  subpoblaciones (distribuidas bajo  $c_i(w; \theta_i)$ ) etiquetadas como  $i = 1, 2, \dots, k$ . Por este motivo introducimos una nueva variable  $z_i \in [1, 2, \dots, k]$  que asigne para cada una de las muestras a cuál de los componentes es más plausible que pertenezca.

Condicionando a esta nueva variable  $z_i$  el modelo de mezclas puede reescribirse como:

$$y_i \mid z_i \sim f_i(y_i \mid \theta_{z_i}, w_{z_i})$$

Notemos que se tienen  $k$  parámetros extras por estimar, que son las proporciones de cada cópula en la función de densidad de la mezcla. Y dado que se tiene una nueva variable  $z_i$  que son datos no observables, es conveniente presentar un algoritmo que trabaje con este tipo de problemas.



## 8.1. Algoritmo EM

Fue propuesto de manera formal por Dempster, Laird, y Rubin (1977) [9], quienes presentaron una técnica iterativa general para realizar una estimación de máxima verosimilitud de parámetros de problemas en los que existe información faltante o no observada.

El algoritmo EM puede aplicarse a muchas situaciones en las que se desea estimar un conjunto de parámetros  $\theta$  que describen una distribución de probabilidad subyacente, dada únicamente una parte observada de los datos completos producidos por la distribución.

En general, supongamos que en cada realización del experimento aleatorio se observa un dato  $y_i$  y existe un parámetro oculto  $z_i$ . Denotamos por  $Y = \{y_1, y_2, \dots, y_n\}$  al conjunto de datos observados en  $n$  realizaciones del experimento, por  $Z = \{z_1, z_2, \dots, z_n\}$  al conjunto de datos no observados y por  $X = Y \cup Z$

Entonces de manera general, el algoritmo EM (Expectation-Maximization), es el siguiente:

---

### Algoritmo 14 Expectation Maximization

---

- 1: Son asignados, mediante un método razonable, los valores iniciales de los parámetros.
  - 2: Los valores perdidos o ausentes son reemplazados por sus valores esperados condicionados a los datos observados y a los valores actuales de los parámetros.
  - 3: **mientras** no se cumpla condición de paro **hacer**
  - 4: Los parámetros son estimados usando los datos observados y los valores asignados a los datos perdidos.
  - 5: Los valores perdidos son calculados asumiendo que los parámetros actuales son correctos.
  - 6: Los parámetros son nuevamente estimados como en el paso 4
  - 7: **fin mientras**
- 

Si se desea mayor información sobre el Algoritmo EM se puede consultar [10, 23].

## 8.2. Algoritmo EM en mezclas de Cópulas Gaussianas

Consideremos una mezcla de dos cópulas Gaussianas,

$$\begin{aligned} c^*(w; \Theta) &= \sum_{i=1}^2 \pi_i c_i(w; \theta_i) \\ &= \pi_1 c_1(w; \theta_1) + \pi_2 c_2(w; \theta_2) \end{aligned}$$

donde la densidad de la  $k$ -ésima Cópula Gaussiana está dada por

$$c_k(x, y) = \frac{1}{\sqrt{1 - \theta_k^2}} \exp\left\{ -\frac{x^2 + y^2 - 2\theta_k xy}{2(1 - \theta_k^2)} \right\} \exp\left\{ \frac{x^2 + y^2}{2} \right\}. \quad (8.2)$$

Notemos que la función log-verosimilitud para la mezcla de cópulas queda expresada de la siguiente manera:

$$l(\Theta) = \log(\pi_1 c_1(w; \theta_1) + \pi_2 c_2(w; \theta_2)),$$

ahora deseamos estimar los parámetros de dependencia de cada una de las cópulas,  $\theta_1$  y  $\theta_2$ , así como las proporciones de cada una en la mezcla  $\pi_1$  y  $\pi_2$ , notemos que las proporciones de las cópulas en la mezcla deben cumplir que  $\pi_1 + \pi_2 = 1$ , entonces será suficiente determinar  $\pi_1$  y hacer  $\pi_2 = 1 - \pi_1$ , para ello aplicaremos el algoritmo EM de la siguiente manera:

Consideremos la densidad de la  $k$ -ésima cópula dada en (8.2). Aplicando logaritmo a la función de densidad, se tiene,

$$l(\theta_k) = \log(c_k(x, y)) = -\frac{1}{2} \log(1 - \theta_k^2) - \frac{x^2 + y^2 - 2\theta_k xy}{2(1 - \theta_k^2)} + \frac{x^2 + y^2}{2}$$

Derivando la función  $l(\theta_k)$  se tiene,

$$\begin{aligned}
 S(\theta_k) &= \frac{\partial l(\theta_k)}{\partial \theta_k} \\
 &= -\frac{1}{2} \left\{ \frac{(1 - \theta_k^2)(-2xy) - (x^2 + y^2 - 2\theta_k xy)(-2\theta_k)}{(1 - \theta_k^2)^2} \right\} \\
 &\quad - \frac{1}{2} \left( \frac{1}{1 - \theta_k^2} \right) (-2\theta_k) \\
 &\propto \frac{\theta_k}{1 - \theta_k^2} - \left\{ \frac{-xy + xy\theta_k^2 + x^2\theta_k + y^2\theta_k - 2\theta_k^2 xy}{(1 - \theta_k^2)^2} \right\} \\
 &\propto \frac{\theta_k}{1 - \theta_k^2} - \frac{\theta_k}{(1 - \theta_k^2)^2} (x^2 + y^2) + \frac{xy}{(1 - \theta_k^2)^2} + \frac{xy\theta_k^2}{(1 - \theta_k^2)^2} \\
 &\propto \frac{\theta_k}{1 - \theta_k^2} \left( 1 - \frac{x^2 + y^2}{1 - \theta_k^2} + \frac{xy\theta_k}{1 - \theta_k^2} \right) + \frac{xy}{(1 - \theta_k^2)^2}
 \end{aligned}$$

Ahora veamos cómo queda el algoritmo EM para una Mezcla de dos Cópulas Gaussianas:

1. Comenzamos inicializando el vector de parámetros de las Cópulas  $\theta_k$  y el vector de pesos  $w_k$  y evaluar la log-verosimilitud con 8.3.
2. **Paso E:** Consiste en calcular el valor esperado condicional actual de los datos no observados  $y_j$  dados los valores observados  $x_j$  y los valores actuales de los parámetros  $\Theta^k$ .

$$E[(y_j)_k | x_k, \Theta^t] = P[(y_j)_k = 1 | x_j, \theta_k] = \frac{P[(y_j)_k = 1] f(x_j | (y_j)_k = 1)}{f(x_j; \theta_k)}$$

En términos de cópulas se tiene,

$$\frac{\pi_k^t c_k(\theta_k^t; u_j, v_j)}{f(\Theta^t; u_j, v_j)} = P[k | u_i, v_i]^t$$

3. **Paso M:** Usando  $P[k|u_i, v_i]^t$ , resolver para  $\theta_k$

$$\begin{aligned} \frac{\partial Q(\theta|\theta^t)}{\partial \theta_k} &= \sum_{i=1}^n \frac{\partial}{\partial \theta_k} (\log[c_k(\theta_k; u_i, v_i)] P[k|u_i, v_i]^t) \\ &= \sum_{i=1}^n P[k|u_i, v_i]^t \frac{\partial}{\partial \theta_k} (\log[c_k(\theta_k; u_i, v_i)]) = 0 \\ &= \sum_{i=1}^n P[k|u_i, v_i]^t S(\theta_k) = 0 \end{aligned}$$

y verificar el criterio de paro, si no se cumple, regresar al paso 2.

### 8.3. Resultados

Cuando se desea modelar un conjunto de datos es útil visualizar su comportamiento para tener una mejor idea sobre el modelo que mejor se ajuste a la muestra; en los EDAs es muy difícil visualizar los datos en cada generación para proponer un modelo. Si se pudieran visualizar los datos o si supiéramos a priori que en la muestra se generan clusters como en la Figura 8.1 o que los datos de la muestra presentan cierto patrón como en la Figura 8.2, elegiríamos una mezcla de cópulas gaussiana como el mejor modelo.

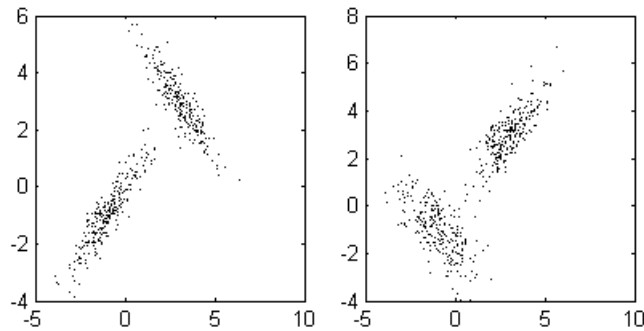


Figura 8.1: Ambas muestras provienen de distribuciones normales bivariadas. (Derecha)  $\mu_1=[-1,-1]$ ,  $\Sigma= [1,0.9; 0.9,1]$ ;  $\mu_2=[3,3]$ ,  $\Sigma= [1,-0.9; -0.9,1]$ ; (Izquierda)  $\mu_1=[-1,-1]$ ,  $\Sigma= [1,-0.6; -0.6,1]$ ;  $\mu_2=[3,3]$ ,  $\Sigma= [1,0.8; 0.8,1]$ ;

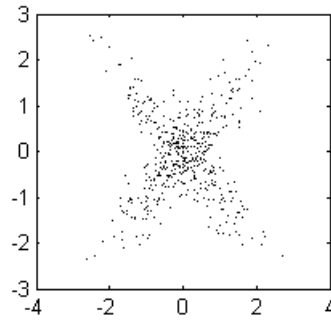


Figura 8.2: Muestra generada de dos distribuciones normales bivariadas con media cero y Matriz de covarianzas  $\Sigma = [1,0.9;0.9,1]$ .

La mezcla de cópulas Gaussianas parece ser una buena alternativa como función de distribución en los EDAs, sin embargo, tienen una serie de inconvenientes con los que nos tenemos que enfrentar. Por ejemplo: determinar el número de cópulas para la mezcla, además de que se desearía poder mezclar diferentes tipos de cópulas. Otro de los problemas es la inicialización correcta de los parámetros, ya que no encontramos una buena estrategia para inicializar el algoritmo EM, por ello, no lo implementamos en un EDA.

En nuestra implantación sólo consideramos cópulas gaussianas estándar, es decir, con media cero y varianza uno, lo que representa una limitante más de éstos modelos, ya que al cambiar la media y la varianza implica incrementar el número de parámetros por estimar.

Las pruebas que se hicieron son para mostrar que el algoritmo EM es una buena opción para estimar los parámetros en este tipo de problemas. Consisten en proponer los parámetros de dependencia  $\Theta = (\theta_1, \theta_2)$  y las proporciones de cada cópula en la mezcla  $\pi_1, \pi_2 = 1 - \pi_1$ . A partir de esa distribución generar una muestra y aplicar el algoritmo EM, para verificar si realmente puede estimar los parámetros que se propusieron.

Se presentan tres de las pruebas realizadas para la estimación de parámetros en una mezcla de dos cópulas gaussianas. En la tabla 8.1 se presentan los parámetros de dependencia para cada cópula  $(\theta_1, \theta_2)$ , los parámetros de la proporción de cada cópula  $(\pi_1, \pi_2)$  y la correlación de la muestra  $(\tau_1, \tau_2)$ , además se anexan los valores que se obtienen con la implementación realizada.

Mezcla de cópulas Gaussianas									
Parámetros Iniciales					Parámetros finales				
$\pi_1$	$\pi_2$	$\theta_1$	$\theta_2$	$\tau$	$\pi_1$	$\pi_2$	$\theta_1$	$\theta_2$	$\tau$
0.5	0.5	0.9	-0.8	0.0857	0.4952	0.5048	0.8908	-0.7841	0.0544
0.2	0.8	0.5	-0.5	-0.2233	0.1984	0.8016	0.3516	-0.4807	0.4031
0.2	0.8	0.9	-0.9	-0.4080	0.1956	0.8044	0.9120	-0.9009	0.4456

Tabla 8.1: Estimación de parámetros para una mezcla de cópulas Gaussianas.

Si observamos la tabla 8.1, notaremos que los parámetros que estimamos con el algoritmo EM son aceptables, y que además, al simular nuevamente con la mezcla de cópulas Gaussianas logramos mantener la correlación entre los datos. En las siguientes figuras se presentan las muestras originales (izquierda) y la nueva muestra que se obtiene con los parámetros estimados (derecha).

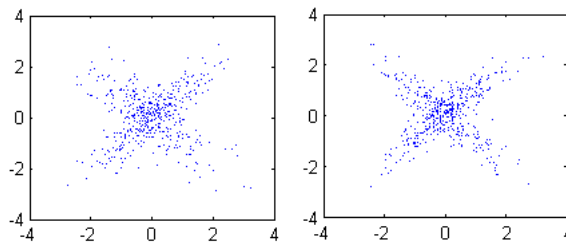


Figura 8.3: Prueba 1. Simulación de una mezcla de cópulas Gaussianas para  $\theta_1 = 0.9$ ,  $\theta_2 = -0.8$ , del lado izquierdo se tiene la muestra original y en el lado derecho la nueva muestra con los parámetros estimados.

En la Figura 8.4 se tiene la misma muestra que en la Figura 8.3, pero después de hacer la transformación al espacio de las cópulas. Notemos que su comportamiento es similar.

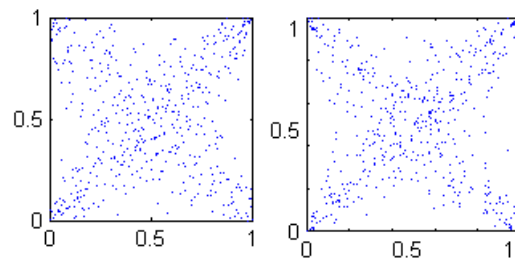


Figura 8.4: Transformación de la muestra, generada en la prueba 1, al espacio de las cópulas.

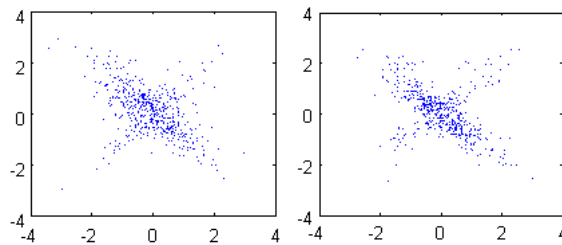


Figura 8.5: Prueba 2. Simulación de una mezcla de cópulas Gaussianas para  $\theta_1 = 0.5$ ,  $\theta_2 = -0.5$ , con diferentes proporciones en la muestra  $\phi_1 = 0,2$ .

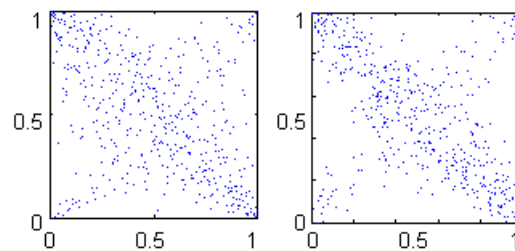


Figura 8.6: Transformación de la muestra generada en la prueba 2 (Figura 8.5).

# Capítulo 9

## Conclusiones

Con éste trabajo de tesis hemos logrado dar un paso importante en el desarrollo de los EDAs, se ha hecho una nueva propuesta para trabajar y explotar su eficiencia frente a problemas de optimización proponiendo una alternativa para modelar los datos de la muestra, que permite capturar y sobre todo, mantener las relaciones que se dan entre los mejores individuos de cada generación.

Es cierto que el estudio de las funciones de cópula merece un espacio mayor al que le hemos dedicado, pero la teoría que existe es muy extensa que sólo nos enfocamos a los aspectos que consideramos de mayor relevancia con lo que logramos otra de las metas, que fue la comprensión y aplicación de las funciones de cópula en EDAs que, como ya lo hemos mencionado, permiten trabajar la función de distribución conjunta de las variables en términos de las distribuciones marginales de cada variable, lo que permite mantener las dependencias capturadas.

Las conclusiones más importantes a las que llegamos después de haber trabajado con funciones de cópula en los EDAs son las siguientes:

- La correlación por rangos es un buen indicador para las relaciones entre variables. El hecho de estar acotada permite tener un patrón de referencia en cuanto a qué tan fuertes son las relaciones que existen entre variables, lo cual sería muy difícil para otras medidas cuyo dominio no es finito. Además, la correlación por rangos permite hallar una relación directa con el parámetro de dependencia de las cópulas, lo que facilita la estimación de dicho parámetro.
- El MIMIC, aunque es uno de los algoritmos con menor complejidad,



ya que solo captura dependencias bivariadas, es una buena elección dentro de los EDAs por no tener un alto grado de complejidad en su construcción y que al combinarlo con funciones de cópula mejora el manejo de las interrelaciones que existen.

- De los resultados numéricos que se obtuvieron, podemos concluir que la cópula Gaussiana es una muy buena elección para modelar datos en los EDAs, ya que reporta, en general, mejores resultados que con selección de cópula.
- El MIMIC-SC implica un mayor costo computacional, por la construcción de la cópula empírica y la evaluación de la muestra en todas las cópulas propuestas, y en muchas ocasiones no conviene la inversión.
- Con el uso de las funciones de cópula en los EDAs, logramos aprovechar las dependencias y así mejoramos el funcionamiento de éstos algoritmos para problemas de optimización.
- Aunque no trabajamos la mezcla de cópulas Gaussianas en los EDAs, pudimos obtener buenos resultados en cuanto a la estimación de parámetros, lo que es un buen avance para trabajos futuros.

# Bibliografía

- [1] Abdelmalik Moujahid, Iñaki Inza, Pedro Larrañaga (1997) *Algoritmos de Estimación de Distribuciones*.
- [2] Ané, T. and C. Kharoubi (2003). *Dependence structure and risk measure*. Journal of Business.
- [3] Arnold, Helen and Shevchenko, Pavel (2006). *Dependence Modelling via the Copula Method prepared*
- [4] Bernoulli, D. (1777). *The most probable choice between several discrepant observations and the formation therefrom of the most likely induction*. Acta Acad. Petrop. 3-33. English translation in *Biometrika*, 48, 3-13 (1962).
- [5] Bosman, P.A.N. and Thierens, D. (1999). *An algorithmic framework for density estimation based evolutionary algorithms*. Utrecht, The Netherlands: Utrecht University: Information and Computing Sciences
- [6] Cherubini, Umberto, Luciano Elisa, Vecchiato Walter (2004) *Copula methods in finance*. Wiley Finance Series.
- [7] De Bonete Jeremy S., Isbell Charles L., Jr. Paul Viola (1997) *MI-MIC: Finding Optima by Estimating Probability Densities*.
- [8] Deheuvels, P. (1979). *La fonction de dépendance empirique et ses propriétés. Un test non paramétrique d'indépendance*. Acad. Roy. Belg. Bull. Cl. Sci.
- [9] Dempster, A. P., Laird, N. M. and Rubin, D. B. (1977). *Maximum likelihood from incomplete data via the EM algorithm*. Journal of the Royal Statistical Society (39), 1 – 38.

- [10] Díaz-Francés, E. *The EM algorithm. An application to Finite Mixture Distributions*. Technical Report D-95-18, Centro de Investigación en Matemáticas, A.C.
- [11] Drouet Mari Dominique, Kotz Samuel (2001). *Correlated and Dependence*. World Scientific Publishing Company.
- [12] Fisher, R. A. (1921). *On the “Probable Error” of a Coefficient of Correlation Deduced from Small Sample*. *Metron*, V.1, pp.3-32.
- [13] Harik, Georges (1999) *Linkage Learning via Probabilistic Modeling in the ECGA* IlliGAL Technical Report 99010. Illinois Genetic Algorithms Laboratory Department of General Engineering.
- [14] Holland John, Henry (1970), *Hierarchical descriptions, universal spaces, and adaptive systems* Arthur W. Burks, editor.
- [15] Joe, H. (1997) *Multivariate Models and Dependence Concepts*. Monographs on Statistics and Applied, Probability 73. Chapman and Hall.
- [16] Joe, H. and Xu, J.J. (1996). *The Estimation Method of Inference Functions for Margins for Multivariate Models*. Dept. of Statistics University of British Columbia, Tech. Rept. 166.
- [17] Kalbfleisch, J. G. (1979) *Probability and Statistical Inference*. Volume 2: Statistical Inference. 2nd edition. New York Berlin Heidelberg Tokyo: Springer-Verlag.
- [18] Kimeldorf, G. and A. R. Sampson (1975). *Uniform representations of bivariate distributions*. *Communications in Statistics*.
- [19] Kotz, Samuel and L. Banks, David (2005). *Encyclopedia of statistical sciences. Vol. 5* Wiley-Interscience; 2 edition
- [20] Kurowicka Dorota, Cooke Roger (2006) *Uncertainty Analysis with High Dimensional Dependence Modelling*. Wiley; 1 edition.

- [21] Larrañaga P. and Lozano J.A. (2002). *Estimation of Distribution Algorithms. A New Tool for Evolutionary Computation*. Kluwer Academic Publishers.
- [22] Lee, L. (1983). *Generalized econometric models with selectivity*. Econometrica.
- [23] M. Bishop , Christopher (2006). *Pattern Recognition and Machine Learning*. Springer
- [24] Mitchell, Melanie (1998). *An Introduction to Genetic Algorithms* A Bradford Book The MIT Press Cambridge, Massachusetts, London, England Fifth printing, 1999
- [25] Morgenstern, D. (1956). *Einfache Beispiele Zweidimensionaler Verteilungen*. Mitteilungsblatt für Mathematische Statistik.
- [26] Mühlenbein H. and Paab G. (1996). *From recombination of genes to the estimation of distribution I. Binary parameters* In Lecture Notes in Computer Science 1411: Parallel Problem Solving from Nature.
- [27] Murphy Kevin, (1998) *A Brief Introduction to Graphical Models and Bayesian Networks*
- [28] Nelsen, R. B. (2006), *An Introduction to Copulas.*, 2nd edition. New York:Springer.
- [29] Pelikan M. and H. Mühlenbein (1999). *The bivariate marginal distribution algorithm*. Advances in Soft Computing-Engineering Design and Manufacturing.
- [30] Pravin K. Trivedi and David M. Zimmer (2007). *Copula Modeling: An Introduction for Practitioners*. World Scientific Publishing.
- [31] Shannon, Claude E. and Weaver, W. (1959). *The Mathematical Theory of Communication*. University of Illinois Press, Champaign, III.
- [32] [http://es.wikipedia.org/wiki/Divergencia\\_de\\_Kullback-Leibler](http://es.wikipedia.org/wiki/Divergencia_de_Kullback-Leibler)
- [33] [http://en.wikipedia.org/wiki/Copula\\_\(statistics\)](http://en.wikipedia.org/wiki/Copula_(statistics))