



Teoría de cópulas y aplicaciones

Ignacio García Rosa



Teoría de cópulas y aplicaciones

Ignacio García Rosa

Memoria presentada como parte de los requisitos para la obtención del título de Máster en Matemáticas por la Universidad de Sevilla.

Tutorizada por

David Gálvez Ruiz

Índice general

English Abstract	1
1. Conceptos y definiciones básicas	3
1.1. Repaso sobre el cálculo probabilístico multidimensional	3
1.2. Definición y primeros resultados sobre cópulas.	5
1.3. El teorema de Sklar	7
1.4. Propiedades de las cópulas	10
2. Medidas de asociación.	17
2.1. Medidas de concordancia	18
2.1.1. Coeficiente ρ_S de Spearman.	19
2.1.2. Coeficiente τ de Kendall.	20
2.1.3. Coeficiente β de Blomqvist.	24
2.1.4. Coeficiente γ de Gini.	25
2.2. Medidas basadas en criterios de información.	26
2.3. Medidas basadas en distancias de L^p	28
2.4. Dependencia de colas.	30

3. Construcción de cópulas	33
3.1. Cópulas elípticas	33
3.1.1. La cópula gaussiana	34
3.1.2. La cópula t -Student	36
3.2. Cópulas de valor extremo	37
3.3. Cópulas arquimedianas	39
3.3.1. Cópula de Clayton	42
3.3.2. Cópula de Gumbel	43
3.3.3. Cópula de Frank	45
4. Aplicación de cópulas al clustering	47
4.1. Algoritmo EM con cópulas.	47
4.1.1. El caso gaussiano.	50
4.2. Algoritmo jerárquico basado en cópulas	55
Anexos	58
A. Demostración del teorema de Sklar	61
B. La cópula empírica	67

English Abstract

Copulas are valuable tools employed in modeling the dependence between different random variables. Their best utility comes from the well-know *Sklar's Theorem*, which establish, for every random vector, the existence of a copula reflexing its dependence structure. In this work, most copula properties are introduced among some parametric families. Finally, as applications, two copula-based models for clustering are developed.

Las cópulas constituyen una gran herramienta a la hora de modelar la dependencia entre una serie de variables aleatorias. Su gran utilidad viene dada por el célebre *Teorema de Sklar*, que indica que, para cualquier vector aleatorio, existe una cópula que refleja su dependencia. A lo largo de este trabajo se exploran las propiedades de estas estructuras matemáticas, y se presentarán diferentes familias parámétricas. Finalmente se propondrán dos modelos de clustering como aplicación práctica.

1 | Conceptos y definiciones básicas

1.1 Repaso sobre el cálculo probabilístico multidimensional

Se comienza este capítulo recordando algunas definiciones y propiedades básicas de las de los vectores aleatorios, así como de sus funciones de distribución multivariante:

Definición 1.1. Se define el σ -álgebra de Borel en \mathbb{R}^n como el menor σ -álgebra que contiene al conjunto $\mathcal{C} = \{B_1 \times \dots \times B_n, B_i \in \mathcal{B}(\mathbb{R})\}$, es decir:

$$\mathcal{B}(\mathbb{R}^n) = \sigma(\mathcal{C}).$$

Definición 1.2. Sea $(\Omega, \mathcal{A}, \mathcal{P})$ un espacio de probabilidad. Se dice que una aplicación $X : \Omega \rightarrow \mathbb{R}^n$ es un vector aleatorio real si es medible para los σ -álgebras \mathcal{A} y $\mathcal{B}(\mathbb{R}^n)$, esto es:

$$X^{-1}(B) = \{\omega \in \Omega : X(\omega) \in B\} \in \mathcal{A}, \quad \text{para todo } B \in \mathcal{B}(\mathbb{R}^n).$$

En este caso X define una probabilidad \mathcal{P}_X en $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$ dada por

$$P_X(B) = \mathcal{P}(X^{-1}(B)), \quad \text{para todo } B \in \mathcal{B}(\mathbb{R}^n).$$

A P_X se la llama distribución de X , y esta caracteriza al vector aleatorio. En el caso particular en que $n = 1$, la aplicación X se denomina variable aleatoria.

Para las siguientes definiciones y proposiciones siempre se tendrá como marco de trabajo un espacio de probabilidad $(\Omega, \mathcal{A}, \mathcal{P})$.

Proposición 1.1. Una aplicación $X : \Omega \rightarrow \mathbb{R}^n$ es un vector aleatorio real si y sólo si cada una de sus funciones componentes es una variable aleatoria real.

Definición 1.3. Gracias a la proposición anterior, dado un vector aleatorio real $X = (X_1, \dots, X_n)$, cada una de las componentes X_i son variables aleatorias reales. Las distribuciones de estas variables aleatorias se denominan distribuciones marginales univariantes de X .

Definición 1.4. Dado un vector aleatorio real $X = (X_1, \dots, X_n)$ se define su función de distribución asociada $F_X : \mathbb{R}^n \rightarrow \mathbb{R}$ como

$$F_X(\underline{x}) = \mathcal{P}(X_1 \leq x_1, \dots, X_n \leq x_n), \quad \text{para todo } \underline{x} \in \mathbb{R}^n.$$

Las definiciones y resultados que se exponen a continuación pueden consultarse en [1].

Definición 1.5. Para una función $F : \mathbb{R}^n \rightarrow \mathbb{R}$, se define el F -volumen del hipercubo

$$[\underline{a}, \underline{b}] = [a_1, b_1] \times \dots \times [a_n, b_n]$$

como

$$V_F([\underline{a}, \underline{b}]) := \sum_{\underline{v}} \text{sign}(\underline{v}) F(\underline{v}),$$

donde la suma se realiza en los 2^d vértices \underline{v} del hipercubo. Aquí:

$$\text{sign}(\underline{v}) = \begin{cases} 1, & \text{si } v_j = a_j \text{ para un número par de índices} \\ -1, & \text{si } v_j = a_j \text{ para un número impar de índices.} \end{cases}$$

Se introduce a continuación una caracterización de las funciones de distribución multivariantes, que se recoge en el siguiente teorema:

Teorema 1.1. $F : \mathbb{R}^n \rightarrow \mathbb{R}$ es la función de distribución de un vector aleatorio X si, y sólo si, F satisface las siguientes propiedades:

1. F es continua por la derecha en cada una de sus variables.
2. El F -volumen de cualquier hipercubo es positivo, es decir, $V_F([\underline{a}, \underline{b}]) \geq 0$ para todo $[\underline{a}, \underline{b}]$ hipercubo.

3. Se verifican los siguientes límites:

$$\begin{aligned} \lim_{x_1 \rightarrow +\infty, \dots, x_n \rightarrow +\infty} F(x_1, \dots, x_n) &= 1 \\ \lim_{x_j \rightarrow -\infty} F(x_1, \dots, x_n) &= 0, \quad \text{para todo } j \in \{1, \dots, n\}. \end{aligned}$$

Observación 1.1.

Algunos comentarios sobre el teorema anterior son los siguientes:

- El anterior teorema nos da una correspondencia unívoca entre las probabilidades definidas sobre $(\mathbb{R}^n, B(\mathbb{R}^n))$ y las funciones F que verifican 1,2 y 3.
- La distribución de un vector aleatorio determina, y es determinada, por su función de distribución.
- Si F tiene derivadas parciales de orden n -ésimo, condición 2 es equivalente a que $\partial^n F / \partial x_1 \dots \partial x_n \geq 0$. De esta misma condición se deduce que F es creciente en cada una de sus coordenadas.

Además, se enuncia la siguiente proposición, que será de gran utilidad en las siguientes secciones.

Proposición 1.2. Sea F una función de distribución multivariante con marginales F_1, \dots, F_n . Entonces se verifica que, dados $\underline{x} = (x_1, \dots, x_n), \underline{y} = (y_1, \dots, y_n) \in \mathbb{R}^n$:

$$|F(x_1, \dots, x_n) - F(y_1, \dots, y_n)| \leq \sum_{i=1}^n |F_i(x_i) - F_i(y_i)|$$

Demostración. Ver [1], Lemas 6.1.8 y 6.1.9. |

1.2 Definición y primeros resultados sobre cópulas.

Para poner al lector en situación, se hará un breve recordatorio sobre las marginales de una función de distribución:

Sea un vector aleatorio $X = (X_1, \dots, X_n)$ con función de distribución F . Por la *Proposición 1.1*, cada una de las componentes del vector constituye una variable aleatoria, y por tanto cada componente tendrá asociada una función de distribución univariante.

Fijada la componente i -ésima, esta función de distribución vendrá dada por:

$$\begin{aligned}
F_i(x_i) &= \mathcal{P}(X_i \leq x_i) \\
&= \mathcal{P}(X_1 \leq +\infty, \dots, X_i \leq x_i, \dots, X_n \leq +\infty) \\
&= \mathcal{P}\left(X_1 \in \bigcup_{n \geq 0} (-\infty, n], \dots, X_i \in (-\infty, x_i], \dots, X_n \in \bigcup_{n \geq 0} (-\infty, n]\right) \\
&= \mathcal{P}\left((X_1, \dots, X_n) \in \bigcup_{n \geq 0} (-\infty, n] \times \dots \times (-\infty, x_i] \times \dots \times (-\infty, n]\right) \\
&= \lim_{n \rightarrow \infty} \mathcal{P}((X_1, \dots, X_n) \in (-\infty, n] \times \dots \times (-\infty, x_i] \times \dots \times (-\infty, n]) \\
&= \lim_{n \rightarrow \infty} F(n, \dots, x_i, \dots, n) = F(+\infty, \dots, x_i, \dots, +\infty)
\end{aligned}$$

A F_i se la denomina como i -ésima función de distribución marginal.

El punto de partida es el siguiente: Como se acaba de probar, dada una función de distribución multivariante, podemos conseguir cada una de las funciones de distribución marginales univariantes sin más que tomar límites. La idea tras las cópulas es exactamente la contraria, es decir, se busca, dado un conjunto de distribuciones marginales univariantes, poder "reconstruir" la distribución multivariante original. Se introduce así el concepto de cópula:

| Definición 1.6. Una cópula C es una función de distribución multivariante de un vector aleatorio X cuyas marginales univariantes siguen distribuciones $U(0, 1)$.

Para el caso de las cópulas bivariantes existe la siguiente caracterización:

Proposición 1.3. Una función $C : [0, 1] \times [0, 1] \rightarrow [0, 1]$ continua a la derecha es una cópula si, y sólo si, verifica:

1. $C(u_1, 1) = u_1$ y $C(1, u_2) = u_2$.
2. $C(u_1, 0) = C(0, u_2) = 0$.
3. Dados $u_1, u_2, v_1, v_2 \in [0, 1]$ con $u_1 \leq v_1$ y $u_2 \leq v_2$, se verifica:

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

Demostración. Las condiciones 2 y 3 surgen Teorema 1.1, teniendo en cuenta que, para $[\underline{u}, \underline{v}] = [u_1, v_1] \times [u_2, v_2]$, se verifica

$$V_C([\underline{u}, \underline{v}]) = C(v_1, v_2) - C(u_1, v_2) - C(v_1, u_2) + C(u_1, u_2).$$

La condición 1, esta es impuesta para que las distribuciones marginales sean distribuciones $U(0, 1)$. |

1.3 El teorema de Sklar

En esta sección se desarrolla el llamado *Teorema de Sklar*, que es el resultado básico que da sentido al estudio de las cópulas.

Definición 1.7. Sea X un vector aleatorio con función de distribución F , y sean X_1, \dots, X_n las marginales univariantes asociadas con funciones de distribución F_1, \dots, F_n respectivamente. Se dice que una cópula C está asociada con X si, para todo $\underline{x} = (x_1, \dots, x_n) \in \mathbb{R}^n$, se verifica que

$$F(\underline{x}) = C(F_1(x_1), \dots, F_n(x_n)).$$

Ejemplo 1.1. Sean X_1, \dots, X_n variables aleatorias reales independientes con funciones de distribución F_1, \dots, F_n respectivamente, entonces la función de distribución del vector aleatorio $X = (X_1, \dots, X_n)$ viene dada por

$$F(\underline{x}) = \prod_{i=1}^n F_i(x_i), \quad \text{para todo } \underline{x} \in \mathbb{R}^n.$$

Por ello una cópula asociada a X en este caso sería la llamada "cópula de independencia", que viene dada por:

$$C_I(\underline{u}) = \prod_{i=1}^n u_i, \quad \text{para todo } \underline{u} \in [0, 1]^n.$$

De hecho, un conjunto de variables aleatorias X_1, \dots, X_n son independientes si, y sólo si, la cópula C_I está asociada al vector aleatorio $X = (X_1, \dots, X_n)$. (Ver [2] *Corollary 7.1.2*).

Para vectores aleatorios continuos es sencillo probar la existencia y unicidad de cópulas asociadas, dicho resultado se recoge en la siguiente proposición:

Proposición 1.4. Sea X un vector aleatorio con función de distribución F continua, entonces existe una única cópula asociada con X , que viene dada por

$$C(\underline{u}) = F(F_1^{-1}(u_1), \dots, F_n^{-1}(u_n))$$

Demostración. Sea X una variable aleatoria continua con función de distribución F , entonces $F(X) \sim U(0, 1)$. En efecto, para ello se consideran los dos casos posibles:

- Caso 1: F es estrictamente creciente y por tanto su inversa está bien definida. Bastanotar que, para $u \in (0, 1)$, $F(x) \leq u$ si, y sólo si, $x \leq F^{-1}(u)$, y por tanto

$$\mathcal{P}(F(X) \leq u) = \mathcal{P}(X \leq F^{-1}(u)) = F(F^{-1}(u)) = u.$$

- Caso 2: F no es estrictamente creciente, es decir, existe algún un punto $u \in (0, 1)$ de forma que $F^{-1}(u)$ es un intervalo. Sea entonces $I = \{x \in \mathbb{R} : F(x) = u\}$. Por ser F continua, I debe ser un intervalo cerrado, además el hecho de considerar $u < 1$ hace que el intervalo I esté acotado superiormente, y por tanto tendría sentido considerar el punto $u' = \text{máx } I$. De este modo

$$\mathcal{P}(F(X) \leq u) = \mathcal{P}(X \leq F^{-1}(u')) = F(F^{-1}(u')) = u.$$

En ambos casos se llega a que

$$\mathcal{P}(F(X) \leq u) = \begin{cases} 0, & \text{si } u \leq 0 \\ u, & \text{si } u \in (0, 1) \\ 1, & \text{si } u \geq 1, \end{cases}$$

que es la función de distribución de una variable $U(0, 1)$.

Dado ahora el vector aleatorio $X = (X_1, \dots, X_n)$, se considera el vector aleatorio $Y = (F_1(X_1), \dots, F_n(X_n))$. Por lo visto hasta ahora, las marginales univariantes de este vector siguen distribuciones $U(0, 1)$, pues cada una de las distribuciones marginales son continuas por serlo X , y por tanto su función de distribución es una cópula. Además, con razonamiento análogo al anterior:

$$\begin{aligned} F_Y(\underline{u}) &= \mathcal{P}(F_1(X_1) \leq u_1, \dots, F_n(X_n) \leq u_n) = \mathcal{P}(X_1 \leq F^{-1}(u_1), \dots, X_n \leq F^{-1}(u_n)) \\ &= F(F^{-1}(u_1), \dots, F^{-1}(u_n)), \end{aligned}$$

con lo que la cópula dada por dicha función de distribución es asociada con X .

Veamos la unicidad:

Sea C la cópula construida, y sea \tilde{C} otra cópula asociada a X . Entonces, por definición, ambas cópulas coinciden en el conjunto

$$A = \{\underline{z} = (z_1, \dots, z_n) : z_i \in F_i(\mathbb{R}), i = 1, \dots, n\}.$$

Ahora bien, como cada F_i es continua, alcanza sobre \mathbb{R} todos los valores del intervalo $(0, 1)$. Es decir, los conjuntos A y $[0, 1]^n$ coinciden salvo, a lo sumo, un número finito de

puntos, luego, denotando como m la medida de Lebesgue n -dimensional, se cumple que $m(A) = m([0, 1]^n)$. En definitiva, C y \tilde{C} coinciden en casi todo $[0, 1]^n$, y como C es continua por ser composición de funciones continuas, debe ser $C = \tilde{C}$ en todo $[0, 1]^n$. |

Una cuestión lógica sería ver si el resultado anterior puede extenderse a vectores aleatorios con funciones de distribución arbitrarias, no necesariamente continuas. La respuesta a ello es dada por el célebre *Teorema de Sklar*:

| Teorema 1.2 (Sklar). *Sea X un vector aleatorio con función de distribución H y funciones de distribución marginales univariantes F_1, \dots, F_n . Entonces existe una cópula C asociada con X , es decir, tal que se verifica que*

$$H(\underline{x}) = C(F_1(x_1), \dots, F_n(x_n)),$$

para todo $\underline{x} \in \mathbb{R}^n$.

Observación 1.2. En los casos en que el vector X no sea absolutamente continuo no existe garantía de que haya una única cópula asociada. De hecho existen contraejemplos de esto, para ello ver [3].

Una prueba de este teorema, junto con algunas propiedades más de la clase de las cópulas como subconjunto del espacio de las funciones continuas en $[0, 1]^n$, puede encontrarse en el *Anexo A*. Este teorema tiene un recíproco, que es de utilidad a la hora de generar muestras de vectores aleatorios arbitrarios:

| Teorema 1.3 (Recíproco del Teorema de Sklar). *Sea C una cópula y sean F_1, \dots, F_n funciones de distribución univariantes cualesquiera. Si se define $H : \mathbb{R}^n \rightarrow [0, 1]$ como*

$$H(\underline{x}) = C(F_1(x_1), \dots, F_n(x_n)).$$

para todo $\underline{x} \in \mathbb{R}^n$, se verifica que H es una función de distribución multivariante.

Demostración. Basta ver que se satisfacen las propiedades del *Teorema 1.1*. Para más detalles, ver [3]. |

Como ya se adelantó, este teorema permite crear vectores aleatorio con distribuciones marginales arbitrarias que verifiquen una determinada relación de dependencia, pues este hecho es controlado por la cópula escogida.

Ejemplo 1.2. Tomando en el *Teorema 1.3* distribuciones marginales $\mathcal{N}(0, 1)$ y la cópula de independencia C_I , se obtiene la función de distribución de un vector aleatorio $\mathcal{N}_n(0, I_{n \times n})$.

1.4 Propiedades de las cópulas

En esta sección se exponen una serie de propiedades que verifican las cópulas. Una consecuencia inmediata de la *Proposición 1.2* es el hecho de que las cópulas son uniformemente continuas:

Proposición 1.5. Cualquier cópula C es uniformemente continua, es decir, dados dos vectores $\underline{u} = (u_1, \dots, u_n), \underline{v} = (v_1, \dots, v_n) \in \mathbb{R}^n$ se verifica:

$$|C(u_1, \dots, u_n) - C(v_1, \dots, v_n)| \leq \sum_{i=1}^n |u_i - v_i|.$$

Demostración. Basta aplicar la *Proposición 1.2* y tener en cuenta que las funciones de distribución marginales de C pertenecen a variables aleatorias $U(0, 1)$. |

Además, para cualquier cópula dada, se pueden obtener cotas fijas de la manera que sigue. Para más información consultar [3], *Capítulo 3*.

Definición 1.8. Se define la cópula inferior de Fréchet como aquella dada por

$$C_L(u_1, \dots, u_n) = \max \left(\sum_{i=1}^n u_i - n + 1 ; 0 \right).$$

Se define también la cópula superior de Fréchet como aquella dada por

$$C_U(u_1, \dots, u_n) = \min(u_1, \dots, u_n).$$

Observación 1.3. Aunque se denomine a C_L como "cópula inferior", lo cierto es que solo es una cópula para el caso $n = 2$, sin embargo, es posible probar que, fijado $\underline{u}_0 \in [0, 1]^n$, siempre existe una cópula C dependiente de \underline{u}_0 , de forma que $C_L(\underline{u}_0) = C(\underline{u}_0)$. Para una prueba ver [4], *Theorem 2.10.13*.

Por otro lado, C_U siempre es una cópula. Para ello se considera una variable aleatoria uniforme U , y el vector aleatorio dado por $X = (U, \dots, U)$. Claramente las marginales de este vector son variables aleatorias uniformes y además se cumple:

$$\begin{aligned} F_X(u_1, \dots, u_n) &= \mathcal{P}(U \leq u_1, \dots, U \leq u_n) = \mathcal{P}(U \leq \min(u_1, \dots, u_n)) \\ &= \min(u_1, \dots, u_n) = C_U \end{aligned}$$

La cópula C_U representa una dependencia positiva perfecta, mientras que C_L representa una dependencia negativa perfecta.

| Teorema 1.4 (Cotas Fréchet-Hoeffding). Para cualquier función de distribución multivariante F con funciones de distribución marginales F_1, \dots, F_n , se verifican las desigualdades

$$\max \left(\sum_{i=1}^n F_i(x_i) - n + 1; 0 \right) \leq F(x_1, \dots, x_n) \leq \min(F_1(x_1), \dots, F_n(x_n)) \quad (1.1)$$

para todo $(x_1, \dots, x_n) \in \mathbb{R}^n$. En consecuencia, dada una cópula C cualquiera, se verifica

$$C_L(u_1, \dots, u_n) \leq C(u_1, \dots, u_n) \leq C_U(u_1, \dots, u_n)$$

para todo $(u_1, \dots, u_n) \in \mathbb{R}^n$.

Demostración. La segunda parte del teorema se deduce de la primera, pues en el caso de las cópulas las marginales siguen distribuciones uniformes. Se procede a probar (1.1):

1. Para la segunda desigualdad basta ver que, por la monotonía de la probabilidad:

$$F(x_1, \dots, x_n) = \mathcal{P}(X_1 \leq x_1, \dots, X_n \leq x_n) \leq \mathcal{P}(X_i \leq x_i) = F_i(x_i)$$

para $i = 1, \dots, n$, de donde se deduce lo deseado.

2. Se considera ahora la segunda desigualdad. Por definición, es claro que $F(x_1, \dots, x_n) \geq 0$. Por otro lado, por la *Desigualdad de Bonferroni* (consultar, por ejemplo, [5]):

$$\begin{aligned} F(x_1, \dots, x_n) &= \mathcal{P}(X_1 \leq x_1, \dots, X_n \leq x_n) \geq 1 - \sum_{i=1}^n \mathcal{P}(X_i > x_i) \\ &= 1 - \sum_{i=1}^n (1 - \mathcal{P}(X_i \leq x_i)) = 1 - n + \sum_{i=1}^n F_i(x_i). \end{aligned}$$

Proposición 1.6. Sea (X_1, \dots, X_n) un vector aleatorio con distribuciones marginales continuas y cópula asociada C , y sean T_1, \dots, T_n funciones estrictamente crecientes. Entonces el vector aleatorio $(T_1(X_1), T_2(X_2), \dots, T_n(X_n))$ también tiene a C como cópula asociada.

Demostración. Debido a que T_1, \dots, T_n son funciones estrictamente crecientes:

$$\begin{aligned} \mathcal{P}(T_1(X_1) \leq x_1, \dots, T_n(X_n) \leq x_n) &= \mathcal{P}(X_1 \leq T^{-1}(x_1), \dots, X_n \leq T^{-1}(x_n)) \\ &= C(\mathcal{P}(X_1 \leq T^{-1}(x_1)), \dots, \mathcal{P}(X_n \leq T^{-1}(x_n))) \\ &= C(\mathcal{P}(T_1(X_1) \leq x_1), \dots, \mathcal{P}(T_n(X_n) \leq x_n)) \end{aligned}$$

Proposición 1.7. Sea C una cópula, entonces existen las derivadas parciales de primer orden $\partial C / \partial u_j$ en casi todo punto u_j . Además, se verifica que

$$0 \leq \frac{\partial C}{\partial u_j}(u_1, \dots, u_n) \leq 1$$

para todo $(u_1, \dots, u_n) \in \mathbb{R}^n$.

Demostración. Como se exponía en la *Observación 1.1*, al ser C una función de distribución verifica que es creciente en cada una de sus componentes. Del hecho de que las funciones monótonas son derivables en casi todo punto (ver [6], pág 96) se desprende la existencia de las derivadas $\partial C / \partial u_j$. Por otro lado, tomando en la *Proposición 1.3* $x_j = u_j, y_j = u_j + h$ y $x_i = y_i = u_i$ para $i \neq j$, se obtiene:

$$\begin{aligned} 0 \leq \left| \frac{\partial C}{\partial u_j}(u_1, \dots, u_n) \right| &= \lim_{h \rightarrow 0} \left| \frac{C(u_1, \dots, u_j + h, \dots, u_n) - C(u_1, \dots, u_j, \dots, u_n)}{h} \right| \\ &\leq \frac{|u_j + h - u_j|}{|h|} = 1 \end{aligned}$$

Para ver algunas propiedades más de las cópulas, se usan dos teoremas clásicos de teoría de la medida: *El teorema de Radon-Nikodym* y el *Teorema de descomposición de Lebesgue*, los cuales se procede a exponer. Una prueba de los mismos puede encontrarse en [2].

Definición 1.9. Sean dos medidas ν y μ definidas sobre un espacio medible (Ω, \mathcal{A}) .

- Se dice que son singulares la una respecto de la otra si están concentradas sobre conjuntos disjuntos, es decir, si existen $A, B \in \mathcal{A}$ con $A \cap B = \emptyset$ y tales que

$$\nu(E) = \nu(E \cap A), \quad \mu(E) = \mu(E \cap B), \quad \text{para todo } E \in \mathcal{A}.$$

Se denota por $\nu \perp \mu$.

- Por otro lado, se dice que ν es absolutamente continua con respecto de μ si se cumple

$$\mu(E) = 0 \Rightarrow \nu(E) = 0, \quad \text{para todo } E \in \mathcal{A},$$

Se denota por $\nu \ll \mu$.

| Teorema 1.5 (Radon-Nikodym). Sea $(\Omega, \mathcal{A}, \mu)$ un espacio de medida σ -finito, es decir, tal que existe una sucesión $\{A_n\} \subset \mathcal{A}$ verificando

$$\bigcup_{n \in \mathbb{N}} A_n = \Omega, \quad \mu(A_n) < +\infty \quad \text{para todo } n \in \mathbb{N}.$$

Sea también ν una medida finita sobre (Ω, \mathcal{A}) de forma que $\nu \ll \mu$, entonces existe una función integrable $f : \Omega \rightarrow [0, +\infty)$ verificando:

$$\nu(A) = \int_A f \, d\mu, \quad \text{para todo } A \text{ medible.}$$

A f se le llama función de densidad de ν respecto a μ , y es única salvo conjunto de medida μ nula.

| Teorema 1.6 (Teorema de descomposición de Lebesgue). Sea $(\Omega, \mathcal{A}, \mu)$ un espacio de medida σ -finito, y sea ν una medida finita sobre (Ω, \mathcal{A}) . Entonces existe un único par de medidas finitas ν_s, ν_a sobre (Ω, \mathcal{A}) tales que

$$\nu = \nu_s + \nu_a, \quad \nu_s \perp \nu_a, \quad \text{y } \nu_s \ll \mu.$$

Proposición 1.8. Sea X un vector aleatorio de forma que su distribución \mathcal{P}_X es absolutamente continua respecto de la medida de Lebesgue, y sea f la densidad asociada a \mathcal{P}_X por el teorema de Radon-Nikodym. Entonces, la función de distribución F del vector aleatorio cumple que

$$\frac{\partial^n F}{\partial x_1 \dots \partial x_n}(x_1, \dots, x_n) \tag{1.2}$$

existe en casi todo punto de \mathbb{R}^n . Es más, se verifica que

$$\frac{\partial^n F}{\partial x_1 \dots \partial x_n}(x_1, \dots, x_n) = f(x_1, \dots, x_n).$$

en casi todo punto de \mathbb{R}^n .

Aplicando estos resultados a las cópulas, se obtiene el siguiente corolario:

Corolario 1.1. Toda cópula C admite una descomposición en una componente absolutamente continua y otra componente singular. Es decir, podemos escribir

$$C(u_1, \dots, u_n) = A_C(u_1, \dots, u_n) + S_C(u_1, \dots, u_n),$$

siendo

$$A_C(u_1, \dots, u_n) = \int_{-\infty}^{u_n} \dots \int_{-\infty}^{u_1} \frac{\partial^n C}{\partial x_1 \dots \partial x_n}(x_1, \dots, x_n) dx_1 \dots dx_n$$

$$S_C(u_1, \dots, u_n) = C(u_1, \dots, u_n) - A_C(u_1, \dots, u_n).$$

Observación 1.4. Del corolario anterior se deducen las siguientes observaciones:

- Si $C \equiv A_C$, entonces C es absolutamente continua, y su densidad viene dada por

$$\frac{\partial^n C}{\partial u_1 \dots \partial u_n}(u_1, \dots, u_n).$$

Consideremos ahora el caso en el que la cópula C está asociada con un vector aleatorio $X = (X_1, \dots, X_n)$ absolutamente continuo, con función de densidad conjunta f y funciones de densidad marginales f_1, \dots, f_n . En este caso C será absolutamente continua, y derivando en la expresión

$$F(\underline{x}) = C(F_1(x_1), \dots, F_n(x_n)).$$

se obtiene

$$f(x_1, \dots, x_n) = c(F_1(x_1), \dots, F_n(x_n))f_1(x_1), \dots, f_n(x_n),$$

donde c denota la función de densidad de la cópula.

- Si $C \equiv S_C$, es decir, si

$$\frac{\partial^n C}{\partial u_1 \dots \partial u_n}(u_1, \dots, u_n) = 0 \quad \text{en casi todo punto de } \mathbb{R}^n,$$

entonces C es singular.

- En otro caso, C tiene una componente absolutamente continua y una componente singular, y ninguna de estas componentes son cópulas, pues no tienen marginales uniformes en $(0, 1)$.

Ejemplo 1.3. Algunos ejemplos son los siguientes:

- La cópula de independencia es absolutamente continua, y su densidad es

$$c(u_1, \dots, u_n) = 1, \quad \text{para todo } (u_1, \dots, u_n) \in \mathbb{R}^n.$$

- La cópula superior de Fréchet es singular.

Se concluye esta sección con una condición suficiente para comprobar que una cópula es absolutamente continua en el caso bidimensional, cuya prueba puede encontrarse en [7].

Proposición 1.9. Sea C una cópula. Si $\partial C(u, v)/\partial v$ y $\partial^2 C(u, v)/\partial u \partial v$ son continuas en $[0, 1]^2$ y $\partial C(u, v)/\partial u$ existe para todo $u \in (0, 1)$ cuando $v = 0$, entonces $\partial C(u, v)/\partial u$ y $\partial^2 C(u, v)/\partial v \partial u$ existen en $(0, 1)^2$ y $\partial^2 C(u, v)/\partial v \partial u = \partial^2 C(u, v)/\partial u \partial v$.

2 | Medidas de asociación.

En este capítulo se exploran las formas mediante las cuales se pueden usar las cópulas para estudiar los diferentes tipos de dependencia entre variables aleatorias, un primer ejemplo de esto sería el *Ejemplo 1.1*. Un método clásico para estudiar la dependencia entre variables aleatorias en el caso bidimensional viene dado por el estudio del coeficiente de correlación lineal ρ :

| Definición 2.1. Dado un vector aleatorio (X, Y) , se define el coeficiente de correlación lineal ρ de Pearson como

$$\rho_{X,Y} = \frac{\text{Cov}(X, Y)}{\sqrt{\text{Var}(X) \text{Var}(Y)}}.$$

Se verifica que $-1 \leq \rho_{X,Y} \leq 1$, el caso $\rho_{X,Y} = 1$ representa relación lineal positiva perfecta, mientras que caso $\rho_{X,Y} = -1$ representa relación lineal negativa perfecta.

La gran popularidad del coeficiente de correlación lineal se debe a sus buenas propiedades, sin embargo, su uso no es adecuado en situaciones en las que la dependencia multivariante es de tipo más general. Algunos de los problemas que presenta son los siguientes :

- Es posible que alguna de las varianzas del denominador no sea finita, en cuyo caso no estaría bien definido.
- Si dos variables aleatorias X, Y son independientes, entonces $\text{Cov}(X, Y) = 0$, y por tanto $\rho_{X,Y} = 0$. Sin embargo, que $\rho_{X,Y} = 0$ no implica que las variables sean independientes, salvo en el caso gaussiano. Es más, el hecho de que el módulo de $\rho_{X,Y}$ sea pequeño no implica necesariamente baja dependencia.
- El coeficiente de correlación lineal no es invariante bajo transformaciones monótonas.

Es por ello que en la siguiente sección se buscan otras medidas alternativas capaces de adaptarse a problemas más generales. Las siguientes definiciones pueden encontrarse [8].

Definición 2.2. Se define \mathcal{X}_n como el conjunto de los vectores aleatorios dimensión n con distribuciones marginales univariantes continuas.

Definición 2.3. Dadas dos cópulas C y C' se dice que C es más concordante que C' , y se denota como $C' \preceq C$ si, $C'(\underline{u}) \leq C(\underline{u})$ para todo $C'(\underline{u}) \in [0, 1]^n$. Esta relación establece un orden parcial en el conjunto de las cópulas (ver [3]).

Definición 2.4. Una medida de asociación es una aplicación $\mathcal{M} : \mathcal{X}_n \rightarrow [-1, 1]$ que verifica las siguientes propiedades:

1. Invariante respecto permutaciones: Para cualquier permutación π , se verifica

$$\mathcal{M}(X_1, \dots, X_n) = \mathcal{M}(X_{\pi(1)}, \dots, X_{\pi(n)}).$$

2. C_I es la cópula asociada a X si, y sólo si, $\mathcal{M}(X) = 0$.
3. Si $n = 2$, $\mathcal{M}(X) = -1$ si, y sólo si, la cópula asociada a X es C_L , y $\mathcal{M}(X) = 1$ si, y sólo si, la cópula asociada a X es C_U . Para $n > 2$, $\mathcal{M}(X) = 1$ si, y sólo si, la cópula asociada a X es C_U .
4. Si X, X' son dos vectores aleatorios con cópulas asociadas C y C' , y se verifica que $C' \preceq C$, entonces $\mathcal{M}(X') \leq \mathcal{M}(X)$.
5. Continuidad: Sea $\{X_n\}$ es una sucesión de vectores aleatorios de forma que la sucesión de cópulas asociadas $\{C_n\}$ converge puntualmente a C . Entonces, si X es un vector aleatorio asociado a C :

$$\lim_n \mathcal{M}(X_n) = \mathcal{M}(X).$$

2.1 Medidas de concordancia

En esta sección se describe como se pueden generalizar al caso multivariantes algunas de las más conocidas medidas para vectores aleatorios bivariantes, tales como la τ_S de Kendall o la ρ de Spearman. Estas medidas se conocen como *medidas de concordancia*, pues verifican la *propiedad 4* de la *Definición 2.4*, así como las *propiedades 1, 2, 5*. Además son invariantes respecto a transformaciones estrictamente monótonas. Para más detalles, ver [8].

2.1.1 Coeficiente ρ_S de Spearman.

Se introduce el coeficiente ρ_S de Spearman, que fue estudiado por primera vez por Spearman, ver [8], y que es una especie de "normalización" del coeficiente de correlación lineal:

Definición 2.5. Se define el coeficiente ρ_S de Spearman asociado a un vector aleatorio bivalente (X, Y) con funciones de distribución marginales F_1 y F_2 como

$$\rho_S = \frac{\text{Cov}[F_1(X), F_2(Y)]}{\sqrt{\text{Var}(F_1(X)) \text{Var}(F_2(Y))}}.$$

Observación 2.1. Debido a la hipótesis de que X e Y son continuos, tanto $F_1(X)$ como $F_2(X)$ siguen distribuciones $U(0, 1)$, que tienen esperanza $1/2$ y varianza $1/12$. De esta forma

$$\rho_S = \frac{E[F_1(X)F_2(Y)] - (1/2)^2}{1/12}.$$

Ahora bien, por lo mostrado en la prueba de la *Proposición 1.4*, por ser (X, Y) un vector aleatorio continuo y ser C su cópula asociada, se debe cumplir que C es también la función de distribución del vector aleatorio $(F_1(X), F_2(Y))$. Por tanto

$$E[F_1(X)F_2(Y)] = \int_{[0,1]^2} uv \, dC(u, v) = \int_{[0,1]^2} C(u, v) \, du \, dv,$$

con lo cual

$$\rho_S = 12 \int_{[0,1]^2} C(u, v) \, du \, dv - 3.$$

Por otro lado, un sencillo cálculo muestra que

$$\int_{[0,1]^2} C_I(u, v) \, du \, dv = 1/4, \quad \int_{[0,1]^2} C_U(u, v) \, du \, dv = 1/6,$$

por lo que es posible reescribir la anterior expresión como

$$\rho_S = \frac{\int_{[0,1]^2} C(u, v) \, du \, dv - \int_{[0,1]^2} C_I(u, v) \, du \, dv}{\int_{[0,1]^2} C_U(u, v) \, du \, dv - \int_{[0,1]^2} C_I(u, v) \, du \, dv}.$$

en consecuencia, es posible interpretar el coeficiente ρ_S como la diferencia media entre las cópulas C y C_I normalizada a través de la diferencia media entre C_U y C_I . La siguiente proposición proporciona una nueva caracterización para el coeficiente ρ_S , y su prueba puede verse en [3].

Proposición 2.1. Sean (X_1, Y_1) , (X_2, Y_2) y (X_3, Y_3) tres vectores aleatorios independientes e idénticamente distribuidos como el vector aleatorio (X, Y) . Entonces el coeficiente ρ_S de (X, Y) se puede escribir como

$$\rho_S = 3(\mathcal{P}((X_1 - X_2)(Y_1 - Y_3) > 0) - \mathcal{P}((X_1 - X_2)(Y_1 - Y_3) < 0)).$$

La *Observación 2.1* permite generalizar el coeficiente ρ_S al caso multivariante de la siguiente manera:

Definición 2.6. Dado un vector aleatorio X con marginales univariantes continuas, y con cópula asociada C , se define su coeficiente ρ_S de Spearman como

$$\rho_S = \frac{\int_{[0,1]^n} C(\underline{u}) d\underline{u} - \int_{[0,1]^n} C_I(\underline{u}) d\underline{u}}{\int_{[0,1]^n} C_U(\underline{u}) \underline{u} - \int_{[0,1]^n} C_I(\underline{u}) d\underline{u}}.$$

Ejemplo 2.1. Para $\theta \in [-1, 1]$, usando la *Proposición 1.3* es sencillo ver que la función dada por

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

es una cópula.

Aplicando el *Teorema de Fubini*:

$$\begin{aligned} \int \int_{[0,1]^2} C_\theta(u, v) dudv &= \int_0^1 \int_0^1 uv + \theta uv(1 - u)(1 - v) dudv \\ &= \int_0^1 \frac{v}{2} + \frac{\theta}{6} v(1 - v) dv \\ &= \frac{1}{4} + \frac{\theta}{36}. \end{aligned}$$

Por tanto, se verifica en este caso que

$$\rho_S = \frac{\theta}{3}.$$

2.1.2 Coeficiente τ de Kendall.

Se procede a exponer el coeficiente τ de Kendall, que fue introducido por Maurice Kendall en 1938. Para ello es necesaria la siguiente definición:

Definición 2.7. Sea (X, Y) un vector aleatorio y $\{(x_i, y_i)\}_{i=1}^n$ una muestra aleatoria simple del mismo. Dado un par de observaciones $(x_i, y_i), (x_j, y_j)$ se dice que es concordante si $x_i \geq x_j$ y $y_i \geq y_j$, o bien si $x_i \leq x_j$ y $y_i \leq y_j$. Se dice que un par de observaciones es discordante si $x_i \geq x_j$ y $y_i \leq y_j$, o bien si $x_i \leq x_j$ y $y_i \geq y_j$.

Definición 2.8. En el caso bivalente, se define el coeficiente τ de Kendall como la probabilidad de obtener un par de observaciones concordantes menos la probabilidad de obtener un par discordante. Más concretamente, si (X_1, Y_1) y (X_2, Y_2) son dos vectores aleatorios independientes e idénticamente como el vector (X, Y) , se define el coeficiente τ de Kendall asociado a (X, Y) como

$$\tau = \mathcal{P}((X_1 - Y_1)(X_2 - Y_2) > 0) - \mathcal{P}((X_1 - Y_1)(X_2 - Y_2) < 0).$$

El siguiente teorema, cuya demostración puede verse en [4], relaciona el coeficiente τ de Kendall con las cópulas:

Teorema 2.1. Sean (X_1, Y_1) y (X_2, Y_2) vectores aleatorios continuos independientes con funciones de distribución conjuntas H_1 y H_2 respectivamente. Se supone además que sus distribuciones coinciden, es decir, X_1 y X_2 tienen una misma función de distribución F , mientras que Y_1 y Y_2 tienen una misma función de distribución G . Sean C_1 y C_2 las cópulas asociadas a (X_1, Y_1) y (X_2, Y_2) respectivamente, entonces, denotando

$$Q = \mathcal{P}((X_1 - Y_1)(X_2 - Y_2) > 0) - \mathcal{P}((X_1 - Y_1)(X_2 - Y_2) < 0),$$

se verifica que

$$Q = 4 \int \int_{[0,1]^2} C_2(u, v) dC_1(u, v) - 1.$$

Demostración. Por ser variables aleatorias continuas, tenemos que

$$\mathcal{P}((X_1 - Y_1)(X_2 - Y_2) < 0) = 1 - \mathcal{P}((X_1 - Y_1)(X_2 - Y_2) > 0),$$

y por tanto

$$Q = 2\mathcal{P}((X_1 - Y_1)(X_2 - Y_2) > 0) - 1. \quad (2.1)$$

Además se verifica que

$$\mathcal{P}((X_1 - Y_1)(X_2 - Y_2) > 0) = \mathcal{P}(X_1 > X_2, Y_1 > Y_2) + \mathcal{P}(X_1 < X_2, Y_1 < Y_2). \quad (2.2)$$

Por otro lado, usando que $H_1 = C_1(F, G)$, se tiene que

$$\begin{aligned} \mathcal{P}(X_1 > X_2, Y_1 > Y_2) &= \mathcal{P}(X_2 < X_1, Y_2 < Y_1) \\ &= \int \int_{\mathbb{R}^2} \mathcal{P}(X_2 \leq x, Y_2 \leq y) dC_1(F(x), G(y)) \\ &= \int \int_{\mathbb{R}^2} C_2(F(x), G(y)) dC_1(F(x), G(y)), \end{aligned}$$

y aplicando el *Teorema del cambio de variable* con la transformación

$$\begin{cases} u = F(x) \\ v = G(y) \end{cases}$$

se obtiene

$$\mathcal{P}(X_1 > X_2, Y_1 > Y_2) = \int \int_{[0,1]^2} C_2(u, v) dC_1(u, v).$$

De forma similar, como $H_2 = C_2(F, G)$, y usando el mismo cambio de variable:

$$\begin{aligned} \mathcal{P}(X_1 < X_2, Y_1 < Y_2) &= \mathcal{P}(X_2 < X_1, Y_2 < Y_1) \\ &= \int \int_{\mathbb{R}^2} \mathcal{P}(X_2 > x, Y_2 > y) dC_1(F(x), G(y)) \\ &= \int \int_{\mathbb{R}^2} 1 - \mathcal{P}(\{X_2 \leq x\} \cup \{Y_2 \leq y\}) dC_1(F(x), G(y)) \\ &= \int \int_{\mathbb{R}^2} (1 - \mathcal{P}(X_2 \leq x) - \mathcal{P}(Y_2 \leq y) + \mathcal{P}(X_2 \leq x, Y_2 \leq y)) dC_1(F(x), G(y)) \\ &= \int \int_{\mathbb{R}^2} (1 - F(x) - G(y) - C_2(F(x), G(y))) dC_1(F(x), G(y)) \\ &= \int \int_{\mathbb{R}^2} (1 - u - v - C_2(u, v)) dC_1(u, v). \end{aligned}$$

Ahora bien, C_1 es una función de distribución de un cierto vector aleatorio (U, V) , siendo U y V variables aleatorias uniformes en $(0, 1)$, luego

$$\begin{cases} \int \int_{[0,1]^2} u dC_1(u, v) = E(U) = \frac{1}{2} \\ \int \int_{[0,1]^2} v dC_1(u, v) = E(V) = \frac{1}{2}, \end{cases}$$

y por tanto,

$$\begin{aligned} \mathcal{P}(X_1 < X_2, Y_1 < Y_2) &= 1 - \frac{1}{2} - \frac{1}{2} + \int \int_{\mathbb{R}^2} C_2(u, v) dC_1(u, v) \\ &= \frac{1}{2} - \frac{1}{2} + \int \int_{\mathbb{R}^2} C_2(u, v) dC_1(u, v). \end{aligned}$$

Sustituyendo esto en las ecuaciones (2.1) y (2.2) se llega a la conclusión. |

Corolario 2.1. Sea (X, Y) un vector aleatorio continuo con cópula asociada C , entonces

$$\tau = 4 \int \int_{[0,1]^2} C(u, v) dC(u, v) - 1.$$

El siguiente lema permite facilitar los cálculos del coeficiente τ de Kendall:

Lema 2.1. Sean C_1 y C_2 dos cópulas, entonces

$$\int \int_{[0,1]^2} C_1(u, v) dC_2(u, v) = \frac{1}{2} - \int \int_{[0,1]^2} \frac{\partial}{\partial u} C_1(u, v) \frac{\partial}{\partial v} C_2(u, v) du dv.$$

Demostración. Supongamos que C_1 y C_2 son absolutamente continuas. En este caso se cumple que

$$\int \int_{[0,1]^2} C_1(u, v) dC_2(u, v) = \frac{1}{2} - \int_0^1 \int_0^1 C_1(u, v) \frac{\partial^2}{\partial u \partial v} C_2(u, v) du dv.$$

Ahora bien, resolviendo la integral interior por partes, y teniendo las propiedades de la *Proposición 1.3* :

$$\begin{aligned} \int_0^1 C_1(u, v) \frac{\partial^2}{\partial u \partial v} C_2(u, v) du &= C_1(u, v) \frac{\partial}{\partial v} C_2(u, v) \Big|_{u=0}^{u=1} - \int_0^1 \frac{\partial^2}{\partial u \partial u} C_1(u, v) \frac{\partial}{\partial v} C_2(u, v) du \\ &= v - \int_0^1 \frac{\partial^2}{\partial u \partial u} C_1(u, v) \frac{\partial}{\partial v} C_2(u, v) du. \end{aligned}$$

Finalmente integrando entre 0 y 1 respecto de v se obtiene el resultado. Para probar el caso general se aproximan las cópulas C_1 y c_2 por sucesiones de cópulas absolutamente continuas, mediante técnicas similares a las usadas en el *Anexo A*. |

Ejemplo 2.2. Considérese la cópula de independencia $C_I(u, v) = uv$. Aplicando el lema anterior

$$\int \int_{[0,1]^2} C_I(u, v) dC_I(u, v) = \frac{1}{2} - \int \int_{[0,1]^2} uv du dv = \frac{1}{2} - \frac{1}{4} = \frac{1}{4},$$

luego

$$\tau = 4 \cdot \frac{1}{4} - 1 = 0.$$

Se comprueba así que el coeficiente τ de Kendall verifica la *propiedad 2* de la *Definición 2.4*.

Los coeficientes τ de Kendall y ρ_S de Spearman son los más usados en la práctica, sin embargo, para una cópula dada, los valores de ambos coeficientes sobre ella pueden ser relativamente distantes. El siguiente teorema da una idea de como de distantes pueden ser:

| **Teorema 2.2.** Sea (X, Y) un vector aleatorio, y sean τ y ρ_S los coeficientes de Kendall y Spearman asociados a dicho vector. Entonces, se verifica que

$$-1 \leq 3\tau - 2\rho_S \leq 1.$$

Demostración. Ver [4], *Theorem 5.1.10*. |

2.1.3 Coeficiente β de Blomqvist.

El coeficiente de Blomqvist, también llamado coeficiente de correlación medial, fue introducido en 1950, y se define como sigue:

Definición 2.9. Dado un vector aleatorio (X, Y) , su coeficiente β de Blomqvist asociado viene dado por

$$\beta = \mathcal{P}((X - \tilde{x})(Y - \tilde{y}) > 0) - \mathcal{P}((X - \tilde{x})(Y - \tilde{y}) < 0),$$

donde \tilde{x} e \tilde{y} representan las medianas de las variables aleatorias X e Y respectivamente.

Este coeficiente también puede reescribirse usando cópulas en el caso continuo, como se recoge en la siguiente proposición:

Proposición 2.2. Sea (X, Y) un vector aleatorio continuo con funciones de distribución conjunta H y funciones de distribución marginales F y G respectivamente. Sea además C la cópula asociada a dicho vector, entonces, bajo estas condiciones, el coeficiente β de Blomqvist asociado viene dado por

$$\beta = 4C(1/2, 1/2) - 1.$$

Demostración. Con un razonamiento análogo al de la prueba del Teorema 2.1, se obtiene:

$$\begin{aligned} \beta &= \mathcal{P}((X - \tilde{x})(Y - \tilde{y}) > 0) - \mathcal{P}((X - \tilde{x})(Y - \tilde{y}) < 0) \\ &= 2(\mathcal{P}(X < \tilde{x}, Y < \tilde{y}) + \mathcal{P}(X > \tilde{x}, Y > \tilde{y})) - 1 \\ &= 2(H(\tilde{x}, \tilde{y}) + [1 - F(\tilde{x}) - G(\tilde{y}) + H(\tilde{x}, \tilde{y})]) - 1. \end{aligned}$$

Teniendo ahora en cuenta que, por definición $F(\tilde{x}) = G(\tilde{y}) = 1/2$, se llega a que

$$\beta = 4H(\tilde{x}, \tilde{y}) - 1 = 4C(F(\tilde{x}), G(\tilde{y})) - 1 = 4C(1/2, 1/2) - 1.$$

Observación 2.2. Como muestra la Proposición 2.2, el coeficiente β de Blomqvist sólo depende del valor de la cópula asociada en el punto $(1/2, 1/2)$, sin embargo, a menudo ofrece una aproximación precisa a los coeficientes τ de Kendall y ρ_S de Spearman, para más información consultar [4].

Ejemplo 2.3. Dado $\theta \in [-1, 1]$, considérese de nuevo la cópula dada por

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

En este caso, el coeficiente β de Blomqvist viene dado por:

$$\beta = 4C_\theta(1/2, 1/2) - 1 = 4 \cdot \left(\frac{1}{4} + \frac{\theta}{16} \right) - 1 = \frac{\theta}{4}.$$

Por lo visto en el *Ejemplo 2.1*, en este caso el coeficiente ρ_S de Spearman tenía un valor de $\frac{\theta}{3}$. Así, la diferencia en valor absoluto entre ambos coeficientes es a lo sumo de $\frac{1}{12}$ (para $\theta = \pm 1$).

2.1.4 Coeficiente γ de Gini.

A partir de los estudios de Corrado Gini en 1910 sobre el llamado *indice di cograduazione semplice*, surge el llamado coeficiente γ de Gini, definido como:

Definición 2.10. Sea (X, Y) un vector aleatorio, con cópula asociada C y funciones de distribución marginales F y G respectivamente. Consideremos las variables aleatorias $U = F(X)$ y $V = G(Y)$, entonces el coeficiente γ de Gini asociado al vector (X, Y) viene dado por

$$\gamma = 2E(|U + V - 1| - |U - V|),$$

es decir,

$$\gamma = 2 \int \int_{[0,1]^2} (|u + v - 1| - |u - v|) dC(u, v).$$

De hecho, este coeficiente se relaciona con las *cotas Fréchet-Hoeffding* de acuerdo con el siguiente teorema:

Teorema 2.3. Sea (X, Y) un vector aleatorio continuo, con cópula asociada C , entonces se verifica que

$$\gamma = \left(4 \int \int_{[0,1]^2} C_L(u, v) dC(u, v) - 1 \right) + \left(4 \int \int_{[0,1]^2} C_U(u, v) dC(u, v) - 1 \right).$$

Demostración. Del hecho de que se puede reescribir

$$C_U(u, v) = \min \{u, v\} = \frac{1}{2} (u + v - |u - v|),$$

se deduce que

$$4 \int \int_{[0,1]^2} C_U(u, v) dC(u, v) - 1 = 2 \int \int_{[0,1]^2} u + v - |u - v| dC(u, v) - 1.$$

Por otro lado, como cualquier cópula es la función de distribución conjunta de un vector aleatorio con marginales uniformes, se verifica

$$\int \int_{[0,1]^2} u dC(u, v) = \int \int_{[0,1]^2} v dC(u, v) = \frac{1}{2},$$

y por tanto

$$4 \int \int_{[0,1]^2} C_U(u, v) dC(u, v) - 1 = 1 - 2 \int \int_{[0,1]^2} |u - v| dC(u, v).$$

De forma similar, se puede reescribir

$$C_L(u, v) = \text{máx} \{u + v - 1, 0\} = \frac{1}{2}(u + v - 1 + |u + v - 1|),$$

por lo que, de forma análoga

$$4 \int \int_{[0,1]^2} C_L(u, v) dC(u, v) - 1 = 2 \int \int_{[0,1]^2} |u + v - 1| dC(u, v) - 1.$$

2.2 Medidas basadas en criterios de información.

En [10] se introduce la *entropía relativa* como una medida de asociación multivariante basada en la teoría de información. Su definición es la siguiente:

Definición 2.11. Dado un vector aleatorio multivariante continuo $X = (X_1, \dots, X_n)$, con función de densidad f y densidades marginales f_1, \dots, f_n respectivamente, se define su entropía relativa como

$$\delta(X) = \int_{\mathbb{R}^n} \log \left[\frac{f(\underline{x})}{\prod_{i=1}^n f_i(x_i)} \right] f(\underline{x}) d\underline{x}.$$

Observación 2.3. Obsérvese que si las componentes del vector aleatorio son independientes, entonces se verifica

$$f(\underline{x}) = \prod_{i=1}^n f_i(x_i),$$

y por tanto, en este caso, $\delta(X) = 0$.

En esta sección se buscará usar este concepto para crear una medida basada en cópulas, y de forma que no dependa de las distribuciones marginales. La siguiente proposición permite reescribir la *entropía relativa* en función de la cópula del vector.

Proposición 2.3. Sea $X = (X_1, \dots, X_n)$ un vector aleatorio multivariante continuo con función de densidad f y densidades marginales f_1, \dots, f_n respectivamente. Entonces

$$\delta(X) = \int_{[0,1]^n} \log[c(\underline{u})]c(\underline{u}) \, d\underline{u},$$

donde c es la función de densidad de la cópula asociada a X .

Demostración. Por lo visto en la *Observación 1.4*, por ser X continuo se cumple

$$f(x_1, \dots, x_n) = c(F_1(x_1), \dots, F_n(x_n))f_1(x_1)\dots f_n(x_n),$$

donde c es la función de densidad de la cópula asociada a X y F_1, \dots, F_n las funciones de distribución marginales. Sustituyendo lo anterior en la expresión de la entropía relativa:

$$\delta(X) = \int_{\mathbb{R}^n} \log [c(F_1(x_1), \dots, F_n(x_n))] c(F_1(x_1), \dots, F_n(x_n))f_1(x_1)\dots f_n(x_n) \, d\underline{x}.$$

Aplicando ahora el *Teorema del cambio de variable* con la transformación $u_i = F_i(x_i)$ para $i = 1, \dots, n$; se obtiene que

$$\delta(X) = \int_{[0,1]^n} \log[c(u_1, \dots, u_n)]c(u_1, \dots, u_n) \, d\underline{u},$$

pues la matriz diferencial asociada al cambio de variable viene dada por

$$A = \text{diag} \{f_1(x_1), \dots, f_n(x_n)\},$$

y por tanto su jacobiano es precisamente $\prod_{i=1}^n f_i(x_i)$, mientras que, por otro lado, como $F_i(\mathbb{R}) = [0, 1]$ para $i = 1, \dots, n$, se deduce que la integral se evalúa sobre el cubo unidad. |

Observación 2.4. Algunas observaciones sobre la proposición anterior son las siguientes:

- $\delta(X) = 0$ si, y sólo si, $c(\underline{u}) \equiv 1$, lo cual sólo ocurre si la cópula asociada a X es C_I . En resumen, $\delta(X) = 0$ si, y sólo si, sus componentes son independientes, y por tanto se verifica la *Propiedad 1* de la *Definición 2.4*.

- La invarianza de las cópulas bajo transformaciones monótonas da también la invarianza de δ . Usando el *Teorema de la convergencia monótona*, se comprueba que también se verifica la *Propiedad 5* de la *Definición 2.4*.
- Para algunas distribuciones, la entropía relativa puede calcularse de forma analítica. Por ejemplo, para distribuciones gaussianas, la entropía relativa tiene la siguiente expresión:

$$\delta(X) = \frac{-1}{2} \log[|\Sigma|],$$

siendo $|\Sigma|$ el determinante de la matriz de correlaciones de X .

2.3 Medidas basadas en distancias de L^p .

Hoeffding fue el primero en considerar medidas de asociación basadas en las distancias en L^p entre una cópula C y la cópula de independencia, C_I (ver [11]). Su trabajo estaba centrado principalmente en el caso $p = 2$, y fue extendido por Schweizer y Wolff en [12], que introdujeron medidas basadas en los casos $p = 1$ y $p = \infty$. Se introducen a continuación estas medidas, así como algunas de sus propiedades.

Definición 2.12. Se define el coeficiente Φ^2 asociado a una cópula C n -dimensional como

$$\Phi^2(C) = h_2(n) \int_{[0,1]^n} [C(\underline{u}) - C_I(\underline{u})]^2 d\underline{u},$$

donde $h_2(n)$ es el factor de normalización dado por

$$h_2(n) = \left(\int_{[0,1]^n} [C_U(\underline{u}) - C_I(\underline{u})]^2 d\underline{u} \right)^{-1} = \left(\frac{2}{(n+1)(n+2)} - \frac{1}{2^n} \frac{n!}{\prod_{i=1}^n (i + \frac{1}{2})} + \frac{1}{3^n} \right)^{-1}.$$

Observación 2.5. En caso de que $\Phi^2(C) = 0$, se tendría que $C = C_I$ en casi todo $[0, 1]^n$, y por ser ambas funciones continuas se desprendería que $C = C_I$ en todo punto, por lo que Φ^2 verifica la *Propiedad 2* de la *Definición 2.4*. De hecho, puede probarse que esta medida verifica también la *Propiedad 4* y la *Propiedad 5* de esta misma definición, para ello consultar [8]. Además la normalización usada implica que $\Phi^2(C_U) = 1$.

Proposición 2.4. Sea $X = (X_1, \dots, X_n)$ un vector aleatorio, y sea X_{n+1} una variable aleatoria independiente de X . Entonces, definiendo $Y = (X_1, \dots, X_n, X_{n+1})$, se

verifica que

$$\Phi^2(Y) < \Phi^2(X).$$

Demostración. Denotando como C_X y C_Y a las cópulas de X e Y respectivamente, es sencillo comprobar que

$$C_Y(u_1, \dots, u_{n+1}) = C_X(u_1, \dots, u_n)u_{n+1}.$$

Por tanto, en casi todo $[0, 1]^{n+1}$, se verifica que

$$\begin{aligned} |C_Y(u_1, \dots, u_{n+1}) - u_1 \dots u_{n+1}|^2 &= |u_{n+1}|^2 |C_X(u_1, \dots, u_n) - u_1 \dots u_n|^2 \\ &< |C_X(u_1, \dots, u_n) - u_1 \dots u_n|^2. \end{aligned}$$

Finalmente, si se integra en $[0, 1]^n$ en la expresión anterior y se aplica el *Teorema de Fubini*, se obtiene que

$$\begin{aligned} \int_{[0,1]^{n+1}} |C_Y(u_1, \dots, u_{n+1}) - u_1 \dots u_{n+1}|^2 dm_{n+1} &< \int_{[0,1]^{n+1}} |C_X(u_1, \dots, u_n) - u_1 \dots u_n|^2 dm_{n+1} \\ &= \int_{[0,1]^n} |C_X(u_1, \dots, u_n) - u_1 \dots u_n|^2 dm_n, \end{aligned}$$

de donde se deduce el resultado. |

Definición 2.13. Se define el coeficiente σ asociado a una cópula C n -dimensional como

$$\sigma(C) = h_1(n) \int_{[0,1]^n} |C(\underline{u}) - C_I(\underline{u})| d\underline{u},$$

donde $h_1(n)$ es el factor de normalización dado por

$$h_1(n) = \left(\int_{[0,1]^n} |C_U(\underline{u}) - C_I(\underline{u})| d\underline{u} \right)^{-1}.$$

Observación 2.6. De nuevo la medida σ verifica las *propiedades 2, 4 y 5* de la *Definición 2.4*, es más, puede probarse que también se verifica la *propiedad 3*, ver [12].

Proposición 2.5. Sea $X = (X_1, \dots, X_n)$ un vector aleatorio, y sea X_{n+1} una variable aleatoria independiente de X . Entonces, definiendo $Y = (X_1, \dots, X_n, X_{n+1})$, se verifica que

$$\sigma(Y) < \sigma(X).$$

Demostración. La prueba es análoga a la de la *Proposición 2.4*. |

| Definición 2.14. Se define el coeficiente \mathcal{K} asociado a una cópula C n -dimensional como

$$\mathcal{K} = h_\infty(n) \sup_{\underline{u} \in [0,1]^n} |C(\underline{u}) - C_I(\underline{u})|,$$

donde $h_\infty(n)$ es el factor de normalización dado por

$$h_\infty(n) = \left(\sup_{\underline{u} \in [0,1]^n} |C_U(\underline{u}) - C_I(\underline{u})| \right)^{-1}.$$

Observación 2.7. La medida \mathcal{K} verifica las propiedades 2, 4 y 5 de la Definición 2.4.

Proposición 2.6. Sea $X = (X_1, \dots, X_n)$ un vector aleatorio, entonces se verifica que

$$\mathcal{K}(X_1, X_2) \leq \mathcal{K}(X_1, X_2, X_3) \leq \dots \leq \mathcal{K}(X_1, X_2, \dots, X_n).$$

Demostración. Ver [8]. |

2.4 Dependencia de colas.

Esta sección da una breve introducción a las llamadas *medidas de dependencia en las colas*. El concepto de dependencia en colas trata de cuantificar el grado de dependencia entre las componentes de una distribución multivariante teniendo en cuenta que algunas de ellas toman valores extremos. Para el caso bivariante, existe la siguiente definición clásica (consultar [4],[8]) :

| Definición 2.15. Sean X e Y variables aleatorias con funciones de distribución F y G respectivamente. Se define el *parámetro de dependencia de colas superior*, λ_U , como el límite, siempre que exista, dado por:

$$\lambda_U = \lim_{t \rightarrow 1^-} \mathcal{P}[Y > G^{-1}(t) \mid X > F^{-1}(t)].$$

De forma similar, se define el *parámetro de dependencia de colas inferior*, λ_L , como el límite, siempre que exista, dado por:

$$\lambda_L = \lim_{t \rightarrow 0^+} \mathcal{P}[Y \leq G^{-1}(t) \mid X \leq F^{-1}(t)].$$

Si $\lambda_U \in (0, 1]$, se dirá que (X, Y) tiene *dependencia de colas superior*, mientras que si $\lambda_U = 0$ se dirá que no hay *dependencia de colas superior*, y exactamente lo mismo para λ_L .

El siguiente teorema muestra que ambos parámetros dependen únicamente de la cópula asociada a (X, Y) :

| Teorema 2.4. Sean X, Y, F, G, λ_U y λ_L tales como en la definición anterior, suponiendo que ambos límites existen. Sea además C una cópula asociada al vector aleatorio (X, Y) , entonces:

$$\lambda_U = \lim_{t \rightarrow 1^-} \frac{1 - 2t + C(t, t)}{1 - t},$$

$$\lambda_L = \lim_{t \rightarrow 0^+} \frac{C(t, t)}{t}.$$

Demostración. Se muestra la prueba para λ_U , ya que para λ_L es muy similar:

$$\begin{aligned} \lambda_U &= \lim_{t \rightarrow 1^-} \mathcal{P}[Y > G^{-1}(t) \mid X > F^{-1}(t)] \\ &= \lim_{t \rightarrow 1^-} \mathcal{P}[G(Y) > t \mid F(X) > t] \\ &= \lim_{t \rightarrow 1^-} \frac{\mathcal{P}[G(Y) > t, F(X) > t]}{\mathcal{P}[F(X) > t]} \\ &= \lim_{t \rightarrow 1^-} \frac{1 - \mathcal{P}[G(Y) \leq t] - \mathcal{P}[F(X) \leq t] + \mathcal{P}[G(Y) \leq t, F(X) \leq t]}{1 - \mathcal{P}[F(X) \leq t]} \\ &= \lim_{t \rightarrow 1^-} \frac{1 - 2t + C(t, t)}{1 - t}. \end{aligned}$$

donde se ha usado que tanto $F(X)$ como $G(Y)$ siguen distribuciones uniformes en $(0, 1)$. |

Ejemplo 2.4. Considerando la familia paramétrica de cópulas dada por

$$C_\theta(u, v) = uv + \theta uv(1 - u)(1 - v), \quad \theta \in [-1, 1],$$

es fácil comprobar que

$$\lambda_L = \lim_{t \rightarrow 0^+} \frac{C_\theta(t, t)}{t} = \lim_{t \rightarrow 0^+} t + \theta t(1 - t)^2 = 0,$$

mientras que el límite λ_U no existe.

3 | Construcción de cópulas

Una vez vista la utilidad de las cópulas para medir la dependencia entre las componentes de las distribuciones multivariantes, en este capítulo se procede a presentar diferentes familias de cópulas, así como métodos de construcción para generarlas, obteniéndose de esta forma una amplia gama de ejemplos que cubren la mayoría de casos posibles.

3.1 Cópulas elípticas

Se considera aquí una popular clase de cópulas usada a menudo en aplicaciones, las *cópulas elípticas*, entre las que destacan dos casos especialmente: La cópula gaussiana y la cópula t -Student. La principal característica de estas cópulas es que, para el caso bivalente, son simétricas respecto a la diagonal principal de $[0, 1]^2$, es decir, respecto a la recta $u = v$. Las cópulas presentadas en esta sección también verifican simetría respecto a la recta $u = 1 - v$.

| Definición 3.1. *Se dice que un vector aleatorio X sigue una distribución elíptica si es absolutamente continuo y su función de densidad es de la forma*

$$g_X(\underline{x}) = A f[(\underline{x} - \underline{\mu})^t \Sigma^{-1} (\underline{x} - \underline{\mu})],$$

donde $\underline{\mu} \in \mathbb{R}^n$, $\Sigma \in \mathcal{M}_{n \times n}$ representan el vector esperanza y la matriz de covarianzas asociados a X respectivamente, f es una función real y positiva, y A es una constante que depende de Σ .

| Definición 3.2. *Se dice que una cópula C es elíptica si existe un vector aleatorio X con distribución elíptica de modo que C es la única cópula asociada a X .*

3.1.1 La cópula gaussiana

Se dice que una cópula es gaussiana si es la cópula asociada de un vector aleatorio gaussiano. Normalmente la cópula más usada en la práctica es la siguiente: Se parte de un vector aleatorio $Y \sim N_n(\underline{\mu}, \Sigma)$, y estandarizando llegamos a otro vector $X \sim N_n(\underline{0}, \Gamma)$, siendo Γ la matriz de correlaciones de Y . Como el proceso de estandarización involucra sólo operaciones monótonas, la cópula de X coincide con la de Y . Así, si Φ_n denota la función de distribución de X , y Φ la función de distribución de la normal univariante estándar, la cópula viene dada por

$$C(u_1, \dots, u_n; \Gamma) = \Phi_n(\Phi^{-1}(u_1), \dots, \Phi^{-1}(u_n)).$$

Usando la *Observación 1.4*, es fácil comprobar que la densidad de la cópula C viene dada por

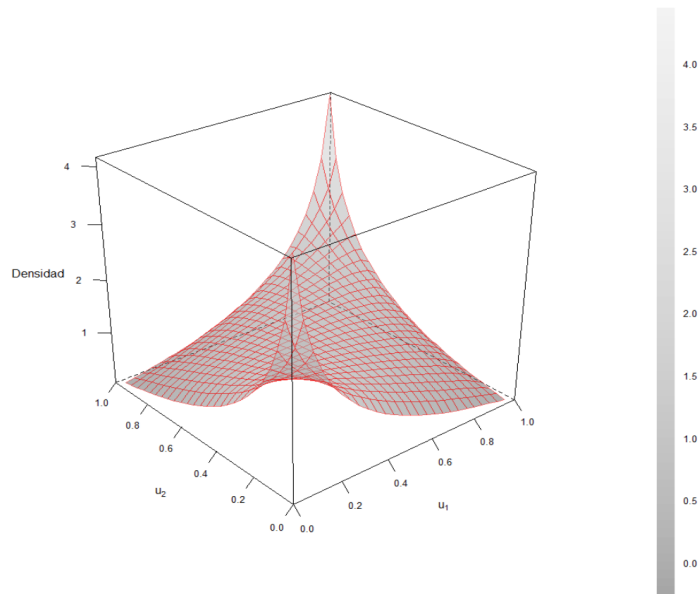
$$c(u_1, \dots, u_n) = |\Gamma|^{-1/2} \exp\left(\frac{-1}{2} \underline{\xi}^t (\Gamma^{-1} - I_n) \underline{\xi}\right),$$

donde $\underline{\xi} = (\xi_1, \dots, \xi_n)$ es el vector de cuantiles, es decir, $\xi_i = \Phi^{-1}(u_i)$ para $i = 1, \dots, n$. En particular, para el caso $n = 2$ la densidad viene dada por

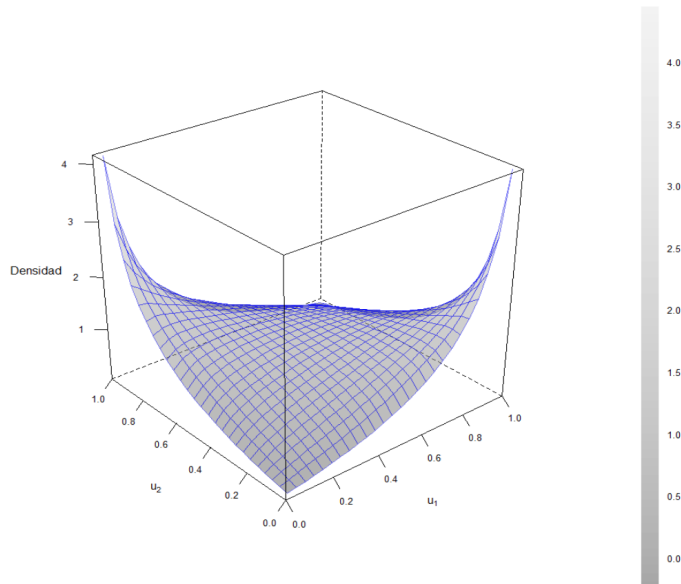
$$c(u_1, u_2; \rho) = \frac{1}{\sqrt{1 - \rho^2}} \exp\left(-\frac{\rho^2 \xi_1^2 - 2\rho \xi_1 \xi_2 + \rho^2 \xi_2^2}{2(1 - \rho^2)}\right),$$

siendo $\rho \in (-1, 1)$ el coeficiente de correlación asociado al vector. Viendo las expresiones anteriores se puede observar que tanto la cópula de independencia (cuando $\Gamma = I_n$) como la cópula C_U (cuando $\Gamma = 1_{n \times n}$) son casos particulares de la cópula gaussiana, es más, para el caso $n = 2$, y tomando $\rho = -1$, se obtiene la cópula C_L . Por ello es posible interpretar la cópula gaussiana como una estructura de dependencia que oscila entre la dependencia perfecta positiva y la dependencia negativa perfecta, indicando el parametro ρ la fuerza de la dependencia.

Densidad de la cópula gaussiana para $\rho = 0.5$:



Densidad de la cópula gaussiana para $\rho = -0.5$:



3.1.2 La cópula t -Student

La cópula t -Student es la cópula asociada a una distribución t -Student multivariante estandarizada $X \sim \mathbf{t}_\nu(\underline{0}, \Gamma)$ con marginales univariantes siguiendo una t -Student univariante t_ν . Denotando por φ_n a la función de distribución de X , y por φ a la función de distribución de la variable aleatoria t_ν , la cópula t -Student de ν grados de libertad viene dada por

$$C(u_1, \dots, u_n; \Gamma) = \varphi_n(\varphi^{-1}(u_1), \dots, \varphi^{-1}(u_n)).$$

Usando de nuevo la *Observación 1.4*, se llega a que la densidad de esta cópula viene dada por

$$c(u_1, \dots, u_n; \Gamma) = K(\nu, n) \frac{1}{|\Gamma|^{-1/2}} [1 + \nu^{-1} \underline{\xi}^t \Gamma^{-1} \underline{\xi}]^{-(\nu+n)/2} \prod_{i=1}^n (1 + \nu^{-1} \xi_i^2)^{(\nu+1)/2},$$

donde $\xi = (\xi_1, \dots, \xi_n)$ es el vector de cuantiles, siendo $\xi_i = \varphi^{-1}(u_i)$ para $i = 1, \dots, n$, y la constante K viene dada por

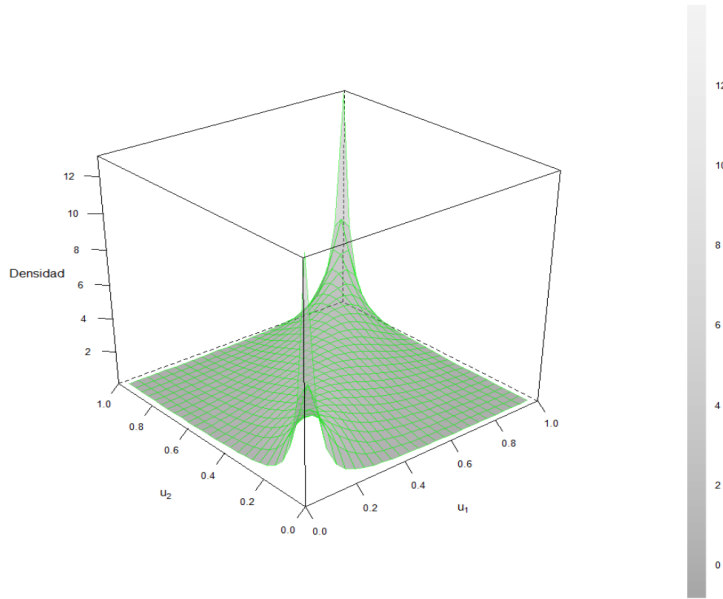
$$K(\nu, n) = f\left(\frac{\nu}{2}\right)^{n-1} f\left(\frac{\nu+1}{2}\right)^{-n} f\left(\frac{\nu+n}{2}\right),$$

con f denotando la *función Gamma*. En particular, para $n = 2$ la densidad de la cópula viene dada por

$$c(u_1, u_2; \rho) = K(\nu, 2) \frac{1}{\sqrt{1-\rho^2}} \left[1 + \frac{\xi_1^2 - 2\rho\xi_1\xi_2 + \xi_2^2}{\nu(1-\rho^2)}\right]^{-(\nu+2)/2} [(1 + \nu^{-1}\xi_1^2)(1 + \nu^{-1}\xi_2^2)]^{(\nu+1)/2},$$

donde $\rho \in (-1, 1)$ es el coeficiente de correlación asociado al vector. Como se puede apreciar, en este caso no se obtiene la cópula de independencia para $\rho = 0$. Esto se debe a que, en el caso de la distribución t de Student, ausencia de correlación no implica independencia, a diferencia del caso gaussiano. Para $\Gamma = 1_{n \times n}$ se vuelve a obtener la cópula C_U .

Densidad de la cópula t de Student para $\rho = 0.8$, $\nu = 4$:



3.2 Cópulas de valor extremo

En esta sección se presentan las *cópulas de valor extremo*, que son especialmente útiles para representar relaciones que ponen mayor énfasis en sucesos extremos de las distribuciones marginales. La forma de definir las es la siguiente:

Para $n \in \mathbb{N}$ dado, sea $X_i = (X_{i,1}, \dots, X_{i,d})$, $i = 1, \dots, n$, una muestra de n vectores aleatorios d -dimensionales continuos, independientes e idénticamente distribuidos con funciones de distribución marginales F_1, \dots, F_d , función de distribución conjunta H y cópula asociada C . Se considera entonces el vector aleatorio de máximos valores en cada variable, dado por

$$M_n = (M_{n,1}, \dots, M_{n,d}), \quad \text{siendo } M_{n,j} = \max_{i=1, \dots, n} \{X_{i,j}\} \quad \text{para } j = 1, \dots, d.$$

Por ser la muestra independiente, las funciones de distribución marginales de M_n vienen dadas por F_1^n, \dots, F_d^n , por tanto, la cópula asociada al vector aleatorio M_n viene dada por

$$C_n(u_1, \dots, u_d) = C(u_1^{1/n}, \dots, u_d^{1/n})^n.$$

En efecto, ya que

$$\begin{aligned} \mathcal{P}(M_{n,1} \leq x_1, \dots, M_{n,d} \leq x_d) &= H(x_1, \dots, x_d)^n \\ &= C(F_1(x_1), \dots, F_d(x_d))^n \\ &= C([F_1(x_1)]^{1/n}, \dots, [F_d(x_d)]^{1/n})^n. \end{aligned}$$

La familia de cópulas de valor extremo surge como el límite de estas cópulas C_n cuando el tamaño muestral n tiende a infinito.

| Definición 3.3. Una cópula C se dice de valor extremo si existe otra cópula C' de manera que

$$C'(u_1^{1/n}, \dots, u_d^{1/n})^n \xrightarrow{n \rightarrow \infty} C(u_1, \dots, u_d).$$

En esta tesis se presenta el siguiente teorema propuesto por Pickands, que da una caracterización de las cópulas de valor extremo en el caso bivalente:

| Teorema 3.1. Una cópula C es una cópula de valor extremo si, y sólo si, existe una función $A : [0, 1] \rightarrow \mathbb{R}$ verificando la siguiente relación:

$$C(u, v) = \exp \left\{ -\log(uv) A \left(\frac{\log(v)}{\log(uv)} \right) \right\}.$$

La función A recibe el nombre de función de dependencia de Pickands asociada a C , y verifica:

1. Es convexa en $[0, 1]$.
2. $\max(t, 1 - t) \leq A(t) \leq 1$, para todo $t \in [0, 1]$.

Demostración. Ver [13]. |

Algunos ejemplos de cópulas de valor extremo son los siguientes:

Ejemplo 3.1. *Cópula de Gumbel:*

Dado $\theta \in [1, +\infty)$, se define la cópula de Gumbel como

$$C_\theta(u, v) = \exp \left\{ - \left([-\log(u)]^\theta + [-\log(v)]^\theta \right)^{1/\theta} \right\}.$$

En este modelo, el parámetro θ mide el grado de dependencia, para $\theta = 1$ tenemos la cópula de independencia, mientras que si $\theta \rightarrow \infty$ obtenemos la cópula C_U . En cuanto a los parámetros de dependencia en colas, se puede comprobar que $\lambda_L = 0$, mientras que $\lambda_U = -2 + 2^{1/\theta}$. La cópula de Gumbel también resulta ser una *cópula arquimediana*, hecho sobre el que se profundizará más en la siguiente sección.

Cópula de Galambos:

Dado $\rho \in [0, +\infty)$, se define la cópula de Galambos como

$$C_\rho(u, v) = \frac{uv}{\exp \left\{ - \left([-\log(u)]^{-1/\rho} + [-\log(v)]^{-1/\rho} \right)^{-\rho} \right\}}.$$

De nuevo el parámetro ρ controla el grado de dependencia, obteniéndose en los casos límites las cópulas C_I y C_U . También se puede comprobar que $\lambda_L = 0$ y $\lambda_U = -2^{-\rho}$.

3.3 Cópulas arquimedianas

A diferencia de las cópulas elípticas, las cuales se obtenían a través de una distribución multivariante dada, en este caso se busca construir cópulas a partir de cierta función generatriz Ψ que verifique ciertas propiedades.

Definición 3.4. Una cópula C es llamada Arquimediana si existe una función continua y estrictamente decreciente $\Psi : [0, 1] \rightarrow [0, +\infty)$ con $\Psi(1) = 0$ de forma que C puede escribirse como

$$C(u_1, \dots, u_n) = \Psi^{[-1]}(\Psi(u_1) + \dots + \Psi(u_n)), \quad (3.1)$$

donde $\Psi^{[-1]}$ es la llamada pseudo inversa de Ψ , dada por

$$\Psi^{[-1]}(u) = \begin{cases} \Psi^{-1}(u), & \text{si } 0 \leq u \leq \Psi(0) \\ 0, & \text{si } \Psi(0) \leq u \leq +\infty. \end{cases}$$

Nótese que $\Psi^{[-1]} = \Psi^{-1}$ cuando se verifica que $\Psi(0) = +\infty$.

Ejemplo 3.2. La cópula de independencia, C_I , es una cópula arquimediana con función generatriz $\Psi(t) = -\log(t)$.

La siguiente proposición recoge algunas propiedades que se desprenden de la propia definición:

Proposición 3.1. Sea C una cópula arquimediana con función generatriz Ψ , entonces:

- C es simétrica, es decir. dado un vector $\underline{u} = (u_1, \dots, u_n) \in \mathbb{R}^n$, se verifica que $C(\underline{u}) = C(\underline{v})$ para cualquier vector \underline{v} resultado de permutar las componentes de \underline{u} .
- Dado $c > 0$ arbitrario, la función $c\Psi$ también es una función generatriz de C .
- Se verifica que C es asociativa, es decir:

$$C(C(u_1, \dots, u_n), u_{n+1}, \dots, u_{2n-1}) = C(u_1, \dots, u_{n-1}, C(u_n, u_{n+1}, \dots, u_{2n-1}))$$

Demostración.

- a) Es consecuencia inmediata de la definición.
- b) Basta observar que $(c\Psi)^{[-1]}(u) = \Psi^{[-1]}(u/c)$.
- c) Se realiza la prueba para el caso $n = 2$, ya que para dimensión mayor la prueba es análoga:

$$\begin{aligned}
C(C(u_1, u_2), u_3) &= \Psi^{[-1]}(\Psi(\Psi^{[-1]}(\Psi(u_1) + \Psi(u_2))) + \Psi(u_3)) \\
&= \Psi^{[-1]}(\Psi(u_1) + \Psi(u_2) + \Psi(u_3)) \\
&= \Psi^{[-1]}(\Psi(u_1) + \Psi(\Psi^{[-1]}(\Psi(u_2) + \Psi(u_3)))) \\
&= C(u_1, C(u_2, u_3)).
\end{aligned}$$

|

Sin embargo, la expresión (3.1) no es una cópula para toda función Ψ . Los dos siguientes teoremas proveen condiciones necesarias y suficientes para que esto ocurra, el primero se centra en el caso bivariante, mientras que el segundo da una caracterización para el caso general. La prueba de ambos teoremas puede encontrarse en [4].

| Teorema 3.2. *Sea una función $\Psi : [0, 1] \rightarrow [0, +\infty)$ continua, estrictamente decreciente y con $\Psi(1) = 0$. Entonces la función*

$$C(u_1, u_2) = \Psi^{[-1]}(\Psi(u_1) + \Psi(u_2))$$

es una cópula si, y sólo si, la función Ψ es convexa.

Para enunciar el caso multivariante es necesaria la siguiente definición:

| Definición 3.5 (Bernstein 1940). *Se dice que una función $g(t)$ es completamente monótona en un intervalo J si $g \in C^\infty(J)$, y además verifica que*

$$(-1)^k \frac{d^k}{dt^k} g(t) \geq 0,$$

para todo $k \in \mathbb{N} \cup \{0\}$, y todo $t \in J$.

| Teorema 3.3. Sea $\Psi : [0, 1] \rightarrow [0, +\infty)$ una función estrictamente decreciente con $\Psi(0) = +\infty$ y $\Psi(1) = 0$, entonces, la función dada por

$$\Psi^{[-1]}(\Psi(u_1) + \dots + \Psi(u_n))$$

es una cópula de dimensión n , para todo $n \geq 2$ si, y sólo si, $\Psi^{[-1]}$ es completamente monótona en $[0, +\infty)$.

Las cópulas arquimedianas son muy útiles, ya que pueden describir de forma sencilla una gran diversidad de estructuras de dependencia. Un ejemplo de ello es el siguiente teorema, que proporciona una manera sencilla de calcular el coeficiente τ de Kendall para este tipo de cópulas:

| Teorema 3.4. Sea (X, Y) un vector aleatorio con cópula asociada C . Si C es una cópula arquimediana con función generatriz Ψ , el coeficiente de Kendall asociado al vector aleatorio viene dado por

$$\tau = 1 + 4 \int_0^1 \frac{\Psi(t)}{\Psi'(t)} dt.$$

Demostración. Ver [4], Corollary 5.1.4. |

Observación 3.1. El Teorema 3.4 proporciona un método para estimar parámetros en el caso de las cópulas arquimedianas:

Sea $(X_1, Y_1), \dots, (X_n, Y_n)$ una muestra aleatoria simple de un vector aleatorio continuo (X, Y) , del que se supone que su cópula asociada C_{ω_0} pertenece a una familia paramétrica de cópulas arquimedianas $\{C_\omega : \omega \in \Omega\}$, donde C_ω viene generada por la función $\Psi_\omega(u)$. Se considera entonces la función $g : \Omega \rightarrow \mathbb{R}$ dada por

$$g(\omega) = 1 + 4 \int_0^1 \frac{\Psi_\omega(t)}{\Psi'_\omega(t)} dt.$$

Suponiendo además que esta función tiene inversa, y en base al Teorema 3.4, un estimador de ω_0 sería $\hat{\omega}_0 = g^{-1}(\hat{\tau})$, donde $\hat{\tau}$ sería un estimador del coeficiente de Kendall de (X, Y) basado en la muestra $(X_1, Y_1), \dots, (X_n, Y_n)$. Para más detalles, ver [23].

Ejemplo 3.3. Para la familia paramétrica de cópulas de Clayton (ver subsección 3.3.1), la función g de la observación anterior viene dada por

$$g(\alpha) = \frac{\alpha}{\alpha + 2}.$$

Por tanto, un estimador para α_0 sería el dado por

$$\hat{\alpha}_0 = \frac{2\hat{\tau}}{1 - \hat{\tau}}.$$

Un resultado similar al *Teorema 3.4* se tiene para los parámetros de dependencia de colas:

Teorema 3.5. *Sea (X, Y) un vector aleatorio con cópula asociada C . Si C es una cópula arquimediana con función generatriz Ψ , se verifica que*

$$\lambda_L = \lim_{t \rightarrow +\infty} \frac{\Psi^{[-1]}(2t)}{\Psi^{[-1]}(t)}$$

$$\lambda_U = 2 - \lim_{t \rightarrow 0^+} \frac{1 - \Psi^{[-1]}(2t)}{1 - \Psi^{[-1]}(t)}.$$

Demostración. Ver [4], Corollary 5.4.3. |

A continuación se exponen algunos ejemplos de familias de cópulas arquimedianas:

3.3.1 Cópula de Clayton

Esta familia paramétrica de cópulas fue estudiada por Kimeldorf y Sampson en 1975 (ver [14]), y viene dada por la función generatriz:

$$\Psi(u) = \alpha^{-1}(u^{-\alpha} - 1), \quad \alpha \in (-1, +\infty) \setminus 0.$$

La función cópula viene dada por

$$C(u_1, \dots, u_n) = \max \{ C(u_1, \dots, u_n) = (u_1^{-\alpha} + \dots + u_n^{-\alpha} - n + 1)^{-1/\alpha}, 0 \},$$

y la densidad de cópula es

$$c(u_1, \dots, u_n) = \left(1 - n + \sum_{i=1}^n u_i^{-\alpha} \right)^{-n-1/\alpha} \prod_{j=1}^n [u_j^{-\alpha-1} (\alpha(j-1) + 1)].$$

Obsérvese que cuando $\alpha \rightarrow +\infty$ se obtiene la cópula C_U , en cambio, si $\alpha \rightarrow 0^+$ se obtiene la cópula C_I . Un inconveniente de esta familia de cópulas es que no puede capturar dependencias negativas.

Para el caso bivalente, es posible calcular el coeficiente τ de Kendall asociado mediante el *Teorema 3.4*, obteniéndose

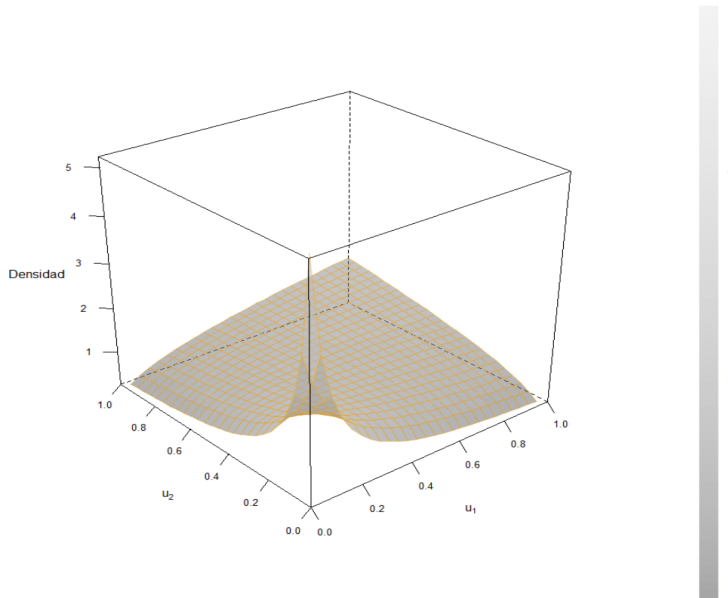
$$\tau = 1 + 4 \int_0^1 \frac{\Psi(t)}{\Psi'(t)} dt = 1 + 4 \int_0^1 \frac{t^{\alpha+1} - t}{\alpha} dt = 1 + \frac{4}{\alpha} \left(\frac{1}{\alpha+2} - \frac{1}{2} \right) = \frac{\alpha}{\alpha+2}.$$

Del mismo modo, usando el *Teorema 3.5*, se obtiene:

$$\lambda_L = \lim_{t \rightarrow +\infty} \frac{\Psi^{[-1]}(2t)}{\Psi^{[-1]}(t)} = \lim_{t \rightarrow +\infty} \frac{(2\alpha t + 1)^{-1/\alpha}}{(\alpha t + 1)^{-1/\alpha}} = 2^{-1/\alpha}$$

$$\lambda_U = 2 - \lim_{t \rightarrow 0^+} \frac{1 - (2\alpha t + 1)^{-1/\alpha}}{1 - (\alpha t + 1)^{-1/\alpha}} = 0.$$

Densidad de la cópula de Clayton para $\alpha = 0.5$:



3.3.2 Cópula de Gumbel

Este tipo de cópulas ya fue introducido en la sección anterior, debido a que son *cópulas de valor extremo*, pero además son cópulas arquimedianas para la función generatriz:

$$\Psi(u) = (-\log u)^\delta, \quad \delta \geq 1.$$

La función cópula viene dada por

$$C(u_1, \dots, u_n) = \exp \left(- \left(\sum_{i=1}^n (-\log u_i)^\delta \right)^{1/\delta} \right).$$

Diferenciando la expresión anterior se obtendría la densidad de cópula. Para el caso $n = 2$ se tiene

$$c(u_1, u_2) = (A + \delta - 1)A^{1-2\delta} \frac{\exp(-A)}{u_1 u_2} (-\log u_1)^{\delta-1} (-\log u_2)^{\delta-1},$$

donde

$$A = [(-\log u_1)^{\delta-1} + (-\log u_2)^{\delta-1}]^{1/\delta}.$$

Se verifica que para $\delta = 1$, la cópula de Gumbel coincide con C_I , y cuando $\delta \rightarrow +\infty$ se obtiene la cópula C_U . Tal y como pasaba con la cópula de Clayton, esta familia tiene el inconveniente de no poder capturar dependencias negativas.

El coeficiente τ de Kendall asociado es

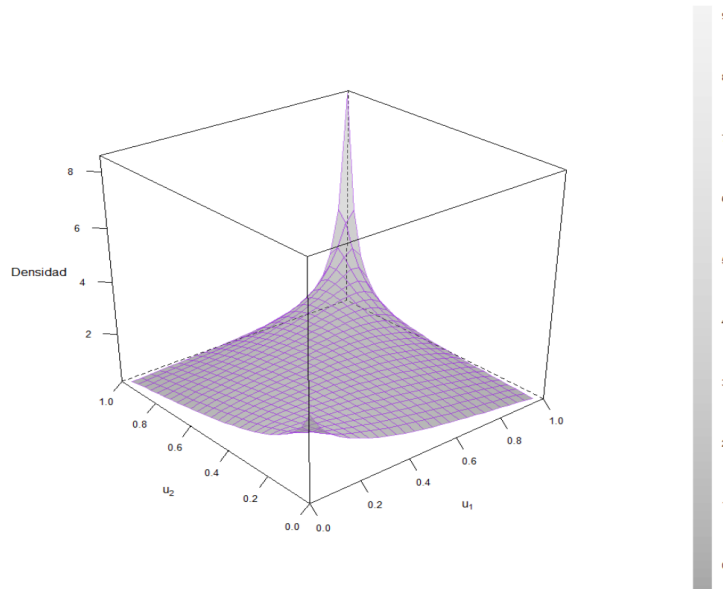
$$\tau = 1 + 4 \int_0^1 \frac{\Psi(t)}{\Psi'(t)} dt = 1 + 4 \int_0^1 \frac{t \log t}{\delta} dt = 1 - \frac{1}{\delta}.$$

En cuanto a los parámetros de dependencia de colas:

$$\lambda_L = \lim_{t \rightarrow +\infty} \frac{\Psi^{[-1]}(2t)}{\Psi^{[-1]}(t)} = \lim_{t \rightarrow +\infty} \frac{\exp(-(2t)^{1/\delta})}{\exp(-t^{1/\delta})} = 0$$

$$\lambda_U = 2 - \lim_{t \rightarrow 0^+} \frac{1 - \exp(-(2t)^{1/\delta})}{1 - \exp(-t^{1/\delta})} = 2 - 2^{1/\delta}.$$

Densidad de la cópula de Gumbel para $\delta = 1.5$:



3.3.3 Cópula de Frank

Esta familia paramétrica de cópulas arquimedianas tiene como función generatriz

$$\Psi(u) = -\log\left(\frac{e^{-\theta u} - 1}{e^{-\theta} - 1}\right), \quad \theta \in \mathbb{R} \setminus \{0\}.$$

La función de cópula viene dada por

$$C(u_1, \dots, u_n) = \frac{-1}{\theta} \log\left(1 + \frac{\prod_{i=1}^n (\exp(-\theta u_i) - 1)}{e^{-\theta} - 1}\right),$$

y para el caso bivalente, la densidad de cópula toma la siguiente expresión:

$$c(u_1, u_2) = \frac{\theta \exp[\theta(1 + u_1 + u_2)](-1 + \exp \theta)}{(\exp \theta - \exp(\theta + \theta u_1) + \exp[\theta(u_1 + u_2)] - \exp(\theta + \theta u_2))^2}.$$

La cópula de Frank no tiene dependencia en colas, y captura dependencias tanto positivas como negativas. De hecho, al igual que la cópula gaussiana, la cópula de Frank puede capturar el rango completo de dependencia, verificandose que:

$$\begin{aligned} \lim_{\theta \rightarrow +\infty} C(u_1, u_2) &= C_U(u_1, u_2) \\ \lim_{\theta \rightarrow -\infty} C(u_1, u_2) &= C_L(u_1, u_2) \\ \lim_{\theta \rightarrow 0} C(u_1, u_2) &= C_I(u_1, u_2). \end{aligned}$$

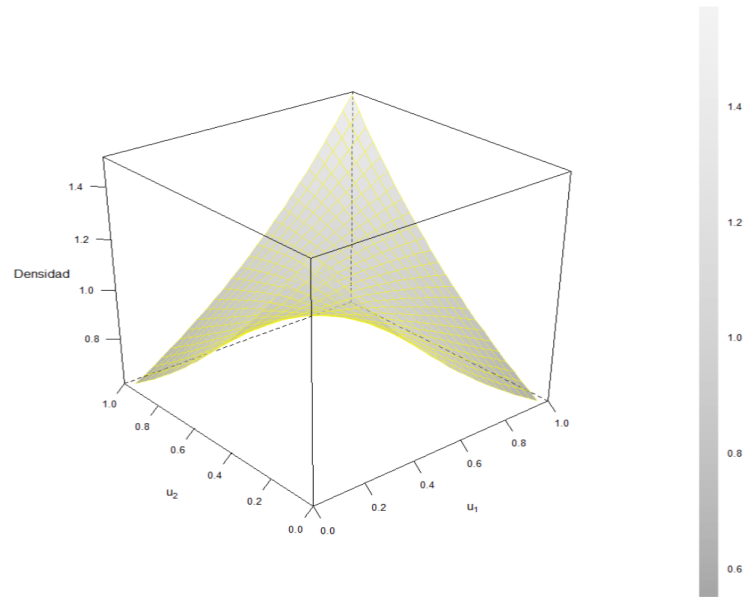
Finalmente, el coeficiente τ de Kendall asociado viene dado por

$$\tau = 1 - 4 \frac{(1 - D_1(\theta))}{\theta},$$

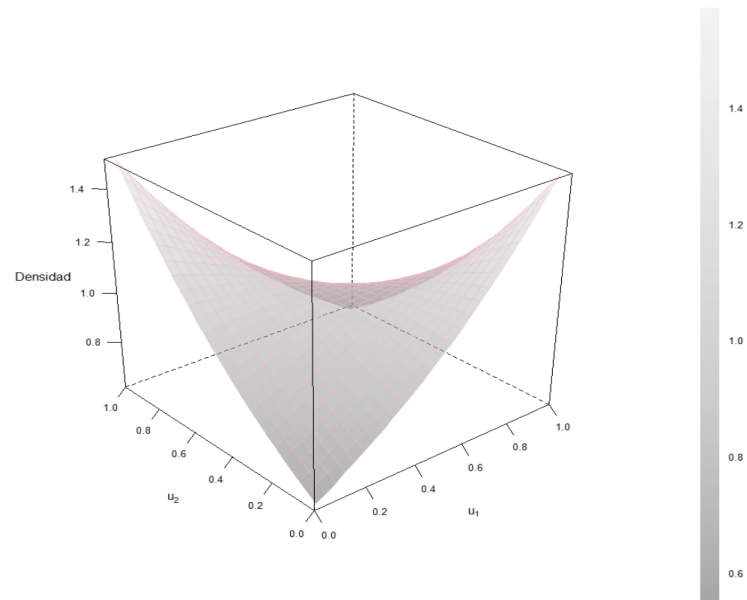
siendo D_k , para $k \in \mathbb{N}$, la k -ésima función de Debye, dada por

$$D_k(\theta) = \frac{k}{x^k} \int_0^\theta \frac{t^k}{e^t - 1} dt.$$

Densidad de la cópula de Frank para $\theta = 1$:



Densidad de la cópula de Frank para $\theta = -1$:



4 | Aplicación de cópulas al clustering

4.1 Algoritmo EM con cópulas.

En esta sección se presenta un algoritmo EM de clustering, ya estudiado en [15], que implementa el uso de las cópulas. La idea reside en clasificar a los individuos de una población teniendo en cuenta la relación existente entre ciertas de sus variables. Además, se tratará en especial el caso gaussiano, el cual puede resolverse de forma analítica.

Sea X una población con n individuos sobre los que puede observarse un conjunto de k variables continuas. El objetivo es establecer g subpoblaciones en X , para ello se supondrá que, para cada una de estas poblaciones, los valores de dichas variables proceden un vector aleatorio cuyas marginales y cópula asociadas pertenecen a familias paramétricas prefijadas. Es deseable que la familia paramétrica de cópulas escogida abarque un gran rango de dependencia.

En los algoritmos EM tradicionales (consultar [16]), se consideran los datos completos $(\underline{Y}, \underline{Z})$, donde \underline{Y} es el vector que recoge las k variables y \underline{Z} es un vector de etiquetas que indica a que subpoblación pertenece el individuo. De esta forma se obtiene la llamada *función log-verosimilitud*, dada por:

$$l(\Psi|\underline{y}, z) = \sum_{j=1}^n \sum_{i=1}^g z_{ij} \log \pi_i + \sum_{j=1}^n \sum_{i=1}^g z_{ij} \log f_i(\underline{y}_j|\underline{\theta}_i),$$

donde $\Psi = (\pi_1, \dots, \pi_g, \underline{\theta}_1, \dots, \underline{\theta}_g)$ es el vector de parámetros. Aquí π_i es la probabilidad a priori de pertenecer a la subpoblación i , y f_i es la función de densidad del vector aleatorio asociado a la subpoblación i , que depende del vector de parámetros $\underline{\theta}_i$. Ahora

bien, haciendo uso de la *Observación 1.4*, es posible expresar

$$f_i(\underline{y}|\theta_i) = c_{\rho_i}(F_{i1}(y_1|\underline{\beta}_{i1}), \dots, F_{ik}(y_k|\underline{\beta}_{ik})) \prod_{m=1}^k g_{im}(y_m|\underline{\beta}_{im}),$$

donde c_{ρ_i} representa la densidad de la cópula C_{ρ_i} , y g_m la densidad de la m -ésima marginal. De esta forma es posible reescribir la *función log-verosimilitud* como

$$\begin{aligned} l(\Psi|\underline{y}, z) &= \sum_{j=1}^n \sum_{i=1}^g z_{ij} \log \pi_i + \sum_{j=1}^n \sum_{i=1}^g z_{ij} \log c_{\rho_i}(F_{i1}(y_{j1}|\underline{\beta}_{i1}), \dots, F_{ik}(y_{jk}|\underline{\beta}_{ik})) \\ &\quad + \sum_{j=1}^n \sum_{i=1}^g z_{ij} \left(\sum_{m=1}^k \log g_{im}(y_{jm}|\underline{\beta}_{im}) \right), \end{aligned}$$

siendo $\Psi = (\pi_1, \dots, \pi_g, \rho_1, \dots, \rho_g, \underline{\beta}_{11}, \dots, \underline{\beta}_{1k}, \dots, \underline{\beta}_{g1}, \dots, \underline{\beta}_{gk})$.

El algoritmo EM se compone de dos pasos en cada iteración, a continuación se expone la iteración $r + 1$ supuesta la iteración r :

1. Paso E:

Sea $\Psi^{(r)}$ la estimación de Ψ correspondiente a la iteración r . El paso E consiste en el cálculo de la esperanza condicional de la función $l(\Psi|\underline{y}, Z)$ dados los datos observados \underline{y} , empleando la estimación $\Psi^{(r)}$. Es decir, se busca calcular $Q(\Psi|\Psi^{(r)}) := E[l(\Psi|\underline{y}, Z)|\underline{Y} = \underline{y}, \Psi^{(r)}]$. De esta forma, se obtiene

$$\begin{aligned} Q(\Psi|\Psi^{(r)}) &= \sum_{j=1}^n \sum_{i=1}^g \tau_{ij}^{(r)} \log \pi_i + \sum_{j=1}^n \sum_{i=1}^g \tau_{ij}^{(r)} \log c_{\rho_i}(F_{i1}(y_{j1}|\underline{\beta}_{i1}), \dots, F_{ik}(y_{jk}|\underline{\beta}_{ik})) \\ &\quad + \sum_{j=1}^n \sum_{i=1}^g \tau_{ij}^{(r)} \left(\sum_{m=1}^k \log g_{im}(y_{jm}|\underline{\beta}_{im}) \right), \end{aligned}$$

siendo $\tau_{ij}^{(r)} = E[Z_{ij}|\underline{Y}_j = \underline{y}_j, \Psi^{(r)}] = P[Z_{ij} = 1|\underline{Y}_j = \underline{y}_j, \Psi^{(r)}]$ la probabilidad de que el individuo j pertenezca al clúster i , que, haciendo uso del *Teorema de Bayes*, puede expresarse como

$$\tau_{ij}^{(r)} = \frac{\hat{\pi}_i^{(r)} c_{\rho_i^{(r)}}(F_{i1}(y_{j1}|\underline{\beta}_{i1}^{(r)}), \dots, F_{ik}(y_{jk}|\underline{\beta}_{ik}^{(r)})) \prod_{m=1}^k g_{im}(y_{jm}|\underline{\beta}_{im}^{(r)})}{\sum_{l=1}^g \hat{\pi}_l^{(r)} c_{\rho_l^{(r)}}(F_{l1}(y_{j1}|\underline{\beta}_{l1}^{(r)}), \dots, F_{lk}(y_{jk}|\underline{\beta}_{lk}^{(r)})) \prod_{m=1}^k g_{lm}(y_{jm}|\underline{\beta}_{lm}^{(r)})}.$$

2. Paso M:

Aquí se maximiza la expresión obtenida en el paso E, estableciéndose así una nueva estimación dada por $\Psi^{(r+1)} = \text{Arg máx}_{\Psi} Q(\Psi|\Psi^{(r)})$.

Nótese que, viendo la expresión de $Q(\Psi|\Psi^{(r)})$, vemos que los parámetros del primer sumando son independientes a los del resto de sumandos, por tanto es posible maximizar dichos parámetros por separado. Para ello, usamos los *Multiplicadores de Lagrange*, obteniéndose

$$\pi^{(r+1)} = \frac{1}{n} \sum_{j=1}^n \tau_{ij}^{(r)}.$$

Para los otros dos sumandos la maximización depende de las familias paramétricas de cópula y marginales escogidas. Estimar simultáneamente los parámetros de cópula y los parámetros marginales es computacionalmente costoso, por tanto, siguiendo [15] y [17], esta maximización se realiza mediante la llamada *máximo verosimilitud en dos pasos*:

En el primer paso, se aplica estimación máxima verosimilitud sólo en el tercer sumando de $Q(\Psi|\Psi^{(r)})$, es decir, en

$$\sum_{j=1}^n \sum_{i=1}^g \tau_{ij}^{(r)} \left(\sum_{m=1}^k \log g_{im}(y_{jm}|\underline{\beta}_{im}) \right).$$

para estimar los parámetros de las marginales. De esta forma se obtienen unos estimadores $\beta_{11}^{(r+1)}, \dots, \beta_{1k}^{(r+1)}, \dots, \beta_{g1}^{(r+1)}, \dots, \beta_{gk}^{(r+1)}$ que verifican

$$(\beta_{11}^{(r+1)}, \dots, \beta_{1k}^{(r+1)}, \dots, \beta_{g1}^{(r+1)}, \dots, \beta_{gk}^{(r+1)}) = \text{Arg máx} \sum_{j=1}^n \sum_{i=1}^g \tau_{ij}^{(r)} \left(\sum_{m=1}^k \log g_{im}(y_{jm}|\underline{\beta}_{im}) \right).$$

En el segundo paso, se estima en parámetro de la cópula usando sólo el segundo sumando de $Q(\Psi|\Psi^{(r)})$ valiéndose de la estimación de los parámetros marginales del paso anterior, es decir, se aplica máxima verosimilitud en

$$\sum_{j=1}^n \sum_{i=1}^g \tau_{ij}^{(r)} \log c_{\rho_i}(F_{i1}(y_{j1}|\hat{\beta}_{i1}), \dots, F_{ik}(y_{jk}|\hat{\beta}_{ik})).$$

De esta forma, se consiguen unos estimadores $\hat{\rho}_1^{(r+1)}, \dots, \hat{\rho}_g^{(r+1)}$ que verifican

$$(\hat{\rho}_1^{(r+1)}, \dots, \hat{\rho}_g^{(r+1)}) = \text{Arg máx} \sum_{j=1}^n \sum_{i=1}^g \tau_{ij}^{(r)} \log c_{\rho_i}(F_{i1}(y_{j1}|\beta_{i1}^{(r+1)}), \dots, F_{ik}(y_{jk}|\beta_{ik}^{(r+1)})).$$

Observación 4.1. Nótese que para aplicar el algoritmo anterior se necesita de una estimación inicial $\Psi^{(0)}$ y un criterio de parada. Para obtener $\Psi^{(0)}$ una forma habitual consiste aplicar cualquier otro algoritmo de clustering, como el k -medias, y estimar en cada subpoblación dada por el algoritmo los parámetros de interés. Como criterio de parada, lo común es detener el algoritmo cuando

$$|l(\Psi^{(t+1)}|\underline{y}, z) - l(\Psi^{(t)}|\underline{y}, z)| < \varepsilon,$$

para un valor prefijado de ε .

Una vez se detiene el algoritmo en el paso r_e , la clasificación de los individuos se realiza de la siguiente manera:

El individuo j -ésimo se asigna a aquella subpoblación i_0 que verifica

$$\tau_{i_0 j}^{(r_e)} = \max \left\{ \tau_{1j}^{(r_e)}, \dots, \tau_{gj}^{(r_e)} \right\},$$

en caso de empate, se asignaría al individuo de forma aleatoria entre aquellas subpoblaciones donde se alcanza el máximo $\tau_{ij}^{(r_e)}$. Para más información sobre estimación de valores iniciales y convergencia del algoritmo, ver [16].

4.1.1 El caso gaussiano.

Aquí se presenta un caso particular del algoritmo anterior, en el cual la maximización de paso M puede realizarse de forma analítica. Se considerará el caso bivalente, con cópula y marginales gaussianas. En este caso, el vector de parámetros viene dado por $\Psi = (\pi_1, \dots, \pi_g, \rho_1, \dots, \rho_g, \underline{\beta}_{11}, \underline{\beta}_{12}, \dots, \underline{\beta}_{g1}, \underline{\beta}_{g2})$, donde $\underline{\beta}_{im} = (\mu_{im}, \sigma_{im}^2)$ para $m = 1, 2$ y $i \in \{1, \dots, g\}$.

Tal como se comentó en el algoritmo general, los parámetros (π_1, \dots, π_g) pueden estimarse de forma independiente, obteniéndose

$$\pi^{(r+1)} = \frac{1}{n} \sum_{j=1}^n \tau_{ij}^{(r)}.$$

De nuevo siguiendo [15] y [17], para los parámetros de cópula y marginales se realiza mediante *máxima verosimilitud en dos pasos*. Se denotarán como $Q_3(\Psi|\Psi^{(r)})$ y $Q_2(\Psi|\Psi^{(r)})$ al tercer y segundo sumando de $Q(\Psi|\Psi^{(r)})$ respectivamente.

- **Estimación de las medias marginales:** Dados $i \in \{1, \dots, g\}$ y $m \in \{1, 2\}$, para obtener $\mu_{im}^{(r+1)}$ se deriva en $Q_3(\Psi|\Psi^{(r)})$ respecto de μ_{im} y se iguala a cero:

$$\begin{aligned} \frac{\partial}{\partial \mu_{im}} Q_3(\Psi|\Psi^{(r)}) &= \sum_{j=1}^n \tau_{ij}^{(r)} \frac{\partial}{\partial \mu_{im}} \log g_m(y_{jm}|\underline{\beta}_{im}) \\ &= \sum_{j=1}^n \tau_{ij}^{(r)} \frac{\partial}{\partial \mu_{im}} \left(-\frac{1}{2} \log(2\pi) - \log \sigma_{im} - \frac{(y_{jm} - \mu_{im})^2}{2\sigma_{im}^2} \right) \\ &= \sum_{j=1}^n \tau_{ij}^{(r)} \left(\frac{y_{jm} - \mu_{im}}{\sigma_{im}^2} \right) = 0. \end{aligned}$$

Despejando en la ecuación anterior se obtiene

$$\mu_{im}^{(r+1)} = \frac{\sum_{j=1}^n \tau_{ij}^{(r)} y_{jm}}{\sum_{j=1}^n \tau_{ij}^{(r)}}.$$

- **Estimación de las varianzas marginales:** Dados $i \in \{1, \dots, g\}$ y $m \in \{1, 2\}$, para obtener $\sigma_{im}^{2(r+1)}$ se deriva en $Q_3(\Psi|\Psi^{(r)})$ respecto de σ_{im} y se iguala a cero:

$$\begin{aligned} \frac{\partial}{\partial \sigma_{im}} Q_3(\Psi|\Psi^{(r)}) &= \sum_{j=1}^n \tau_{ij}^{(r)} \frac{\partial}{\partial \sigma_{im}} \log g_m(y_{jm}|\underline{\beta}_{im}) \\ &= \sum_{j=1}^n \tau_{ij}^{(r)} \frac{\partial}{\partial \sigma_{im}} \left(-\frac{1}{2} \log(2\pi) - \log \sigma_{im} - \frac{(y_{jm} - \mu_{im})^2}{2\sigma_{im}^2} \right) \\ &= \sum_{j=1}^n \tau_{ij}^{(r)} \left[\frac{-\sigma_{im}^2 + (y_{jm} - \mu_{im})^2}{\sigma_{im}^3} \right] = 0. \end{aligned}$$

Despejando en la ecuación anterior se obtiene

$$\sigma_{im}^{2(r+1)} = \frac{\sum_{j=1}^n \tau_{ij}^{(r)} (y_{jm} - \mu_{im})^2}{\sum_{j=1}^n \tau_{ij}^{(r)}},$$

y usando las estimaciones de las medias marginales:

$$\sigma_{im}^{2(r+1)} = \frac{\sum_{j=1}^n \tau_{ij}^{(r)} (y_{jm} - \mu_{im}^{(r+1)})^2}{\sum_{j=1}^n \tau_{ij}^{(r)}}.$$

- **Estimación de la cópula:** Para no sobrecargar las expresiones, se denotará

$$\underline{u}_j^i = (u_{j1}^i, u_{j2}^i) := (F_{i1}(y_{j1}|\underline{\beta}_{i1}^{(r+1)}), F_{i2}(y_{j2}|\underline{\beta}_{i2}^{(r+1)}))$$

para $j \in \{1, \dots, n\}$.

Tal como se vió en el *Capítulo 3*, la densidad de la cópula gaussiana bivalente viene dada por

$$c(u_1, u_2; \rho) = \frac{1}{\sqrt{1 - \rho^2}} \exp\left(-\frac{\rho^2 \xi_1^2 - 2\rho \xi_1 \xi_2 + \rho^2 \xi_2^2}{2(1 - \rho^2)}\right),$$

donde $\xi_m = \Phi^{-1}(u_m)$, para $m = 1, 2$, siendo Φ la función de distribución de la distribución estándar univariante, y $\rho \in (-1, 1)$. El objetivo es maximizar $Q_2(\Psi|\Psi^{(r)})$ usando las estimaciones ya obtenidas para las marginales, es decir, se busca maximizar

$$\Gamma(\underline{\rho}) := \sum_{j=1}^n \sum_{i=1}^g \tau_{ij}^{(r)} \log c_{\rho_i}(u_{j1}^i, u_{j2}^i)$$

respecto ρ_1, \dots, ρ_g . Dado $i \in \{1, \dots, g\}$, derivando respecto ρ_i e igualando a cero:

$$\begin{aligned} \frac{\partial}{\partial \rho_i} \Gamma(\underline{\rho}) &= \sum_{j=1}^n \tau_{ij}^{(r)} \frac{\partial}{\partial \rho_i} \log c_{\rho_i}(u_{j1}^i, u_{j2}^i) \\ &= \sum_{j=1}^n \tau_{ij}^{(r)} \left[\frac{-\rho_i^3 + \xi_{j1}^i \xi_{j2}^i \rho_i^2 + (1 - (\xi_{j1}^i)^2 - (\xi_{j2}^i)^2) \rho_i + \xi_{j1}^i \xi_{j2}^i}{(1 - \rho_i^2)^2} \right] = 0, \end{aligned}$$

Donde $\xi_{jm}^i = \Phi^{-1}(u_{jm}^i)$, para $m = 1, 2$ y $j = 1, \dots, n$. Por tanto, para obtener $\rho_i^{(r+1)}$ basta resolver la ecuación polinómica de grado 3 dada por

$$\begin{aligned} \left(- \sum_{j=1}^n \tau_{ij}^{(r)} \right) \rho_i^3 + \left(\sum_{j=1}^n \tau_{ij}^{(r)} \xi_{j1}^i \xi_{j2}^i \right) \rho_i^2 + \left(\sum_{j=1}^n \tau_{ij}^{(r)} (1 - (\xi_{j1}^i)^2 - (\xi_{j2}^i)^2) \right) \rho_i \\ + \left(\sum_{j=1}^n \tau_{ij}^{(r)} \xi_{j1}^i \xi_{j2}^i \right) = 0. \end{aligned}$$

Ejemplo 4.1. Para probar el algoritmo, se procede a crear una muestra de tamaño $n = 1500$ donde:

1. 500 individuos proceden de una distribución normal multivariante de parámetros

$$\underline{\mu} = \begin{pmatrix} 2 \\ 2 \end{pmatrix} \quad \text{y} \quad \Sigma = \begin{pmatrix} 1 & 0.5 \\ 0.5 & 1 \end{pmatrix}.$$

2. 500 individuos proceden de una distribución normal multivariante de parámetros

$$\underline{\mu} = \begin{pmatrix} -2 \\ -2 \end{pmatrix} \quad \text{y} \quad \Sigma = \begin{pmatrix} 1 & -0.5 \\ -0.5 & 1 \end{pmatrix}.$$

3. 500 individuos proceden de una distribución normal multivariante de parámetros

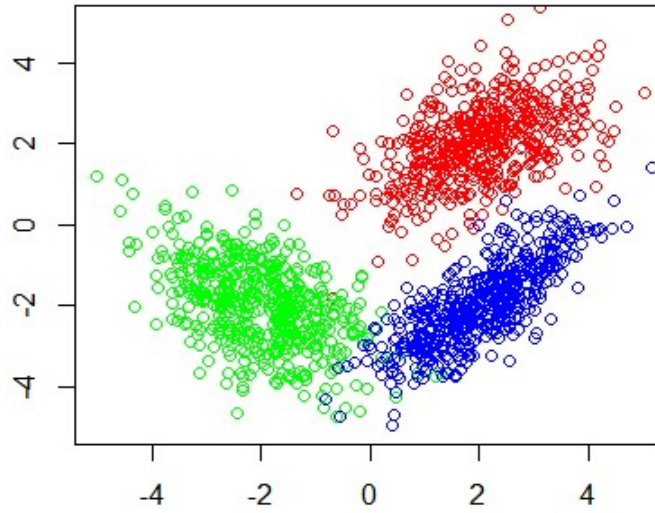
$$\underline{\mu} = \begin{pmatrix} 2 \\ -2 \end{pmatrix} \quad \text{y} \quad \Sigma = \begin{pmatrix} 1 & 0.8 \\ 0.8 & 1 \end{pmatrix}.$$

A continuación se aplican 10 iteraciones del algoritmo con valores iniciales

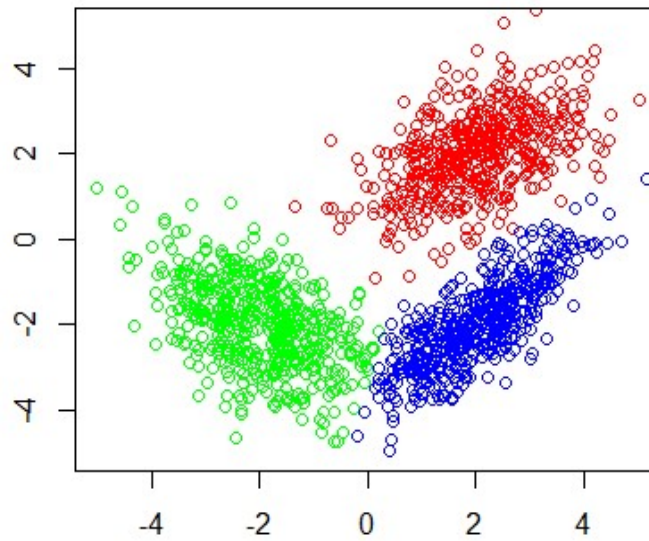
$$\begin{aligned} \underline{\pi}^{(0)} &= (1/3, 1/3, 1/3) \\ \{\mu_{mi}^{(0)}\} &= \begin{pmatrix} 1.5 & -1.5 & 1.5 \\ 1.5 & -1.5 & -1.5 \end{pmatrix} \\ \{\sigma_{mi}^{2(0)}\} &= \begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix} \\ \underline{\rho}^{(0)} &= (0.5, -0.8, 0.8) \end{aligned}$$

dando lugar a los siguientes resultados:

Datos originales:



Datos clasificados por el algoritmo:



4.2 Algoritmo jerárquico basado en cópulas

En esta sección se describe un algoritmo jerárquico de clustering de variables basado en cópulas. En el *Capítulo 3*, se introdujo el coeficiente Φ^2 asociado a una cópula C , el cual se definía como

$$\Phi^2(C) = h_2(d) \int_{[0,1]^d} [C(\underline{u}) - C_I(\underline{u})]^2 d\underline{u},$$

siendo d la dimensión, y donde $h_2(d)$ es el factor de normalización dado por

$$h_2(d) = \left(\int_{[0,1]^d} [C_U(\underline{u}) - C_I(\underline{u})]^2 d\underline{u} \right)^{-1} = \left(\frac{2}{(d+1)(d+2)} - \frac{1}{2^d} \frac{d!}{\prod_{i=1}^d (i + \frac{1}{2})} + \frac{1}{3^d} \right)^{-1}.$$

Usando la *cópula empírica* C_n , (ver anexo B) podemos estimar

$$\int_{[0,1]^d} [C(\underline{u}) - C_I(\underline{u})]^2 d\underline{u}$$

como

$$\begin{aligned} \int_{[0,1]^d} [C_n(\underline{u}) - C_I(\underline{u})]^2 d\underline{u} &= \frac{1}{n^2} \sum_{j=1}^n \sum_{k=1}^n \prod_{i=1}^d (1 - \max\{U_{ji}, U_{ki}\}) \\ &\quad - \frac{2}{n} \left(\frac{1}{2}\right)^d \sum_{j=1}^n \prod_{i=1}^d (1 - U_{ji}^2) + \frac{1}{3^d}, \end{aligned}$$

para más detalles, ver [8].

La idea es crear una medida de similitud inspirada en Φ^2 : si dos clúster de variables están relacionados, la cópula asociada vector aleatorio que recoge las variables de ambos clusters estará lejos de la cópula de independencia, y por el contrario, si las variables de ambos clústers están cerca de ser independientes, la distancia a la cópula de independencia decaerá. Se define por tanto la medida de similitud Δ^2 como sigue:

Definición 4.1. Sean $\mathcal{A}_1 = \{X_1, \dots, X_t\}$ y $\mathcal{A}_2 = \{Y_1, \dots, Y_s\}$ dos clústers de variables. Sea $Z = (X_1, \dots, X_t, Y_1, \dots, Y_s)$ el vector aleatorio formado por las variables aleatorias de ambos clústers, y sea C la cópula asociada a dicho vector. Se define entonces la medida de similitud Δ^2 entre ambos clústers como

$$\Delta^2(\mathcal{A}_1, \mathcal{A}_2) = \frac{10}{h_2(s+t)} \int_{[0,1]^{t+s}} [C(\underline{u}) - C_I(\underline{u})]^2 d\underline{u}.$$

Por lo visto hasta ahora, una estimación natural de dicha medida de similitud sería la dada por

$$\hat{\Delta}^2(\mathcal{A}_1, \mathcal{A}_2) = \frac{10}{h_2(s+t)} \int_{[0,1]^{t+s}} [C_n(\underline{u}) - C_I(\underline{u})]^2 d\underline{u}.$$

El motivo por el cual la constante $h(d)$ se usa dividiendo es que su valor aumenta con la dimensión, y de esta forma se evita que el algoritmo tienda en cada paso a asignar las variables a los clústers con mayor dimensión. Por otro lado, el motivo de multiplicar por 10 es que los valores de la matriz de similitud sean más fáciles de interpretar.

Lo que sigue es incorporar la medida de similitud $\hat{\Delta}^2$ a un algoritmo jerárquico asociativo, tal y como se expone en el siguiente ejemplo:

Ejemplo 4.2. Se crea una muestra de 7 variables aleatorias (700 realizaciones por variable) de la siguiente forma:

1. Las variables X_1, X_2 y X_3 proceden de una cópula de Gumbel 3-dimensional de parámetro $\delta = 5$.
2. Las variables X_4 y X_5 tienen una dependencia exacta: X_4 procede de una distribución $U(0, 1)$ y $X_5 = X_4^2$.
3. Las variables X_6 y X_7 proceden de una cópula gaussiana bivalente de parámetro $\rho = -0.9$.

Se procede con el algoritmo para elaborar 3 clústers: (Nota: Debido a que se está usando una aproximación basada en la cópula empírica, los valores obtenidos a lo largo de la aplicación del algoritmo no son exactos) :

- **Paso 1 :** Se consideran 7 clústers, cada uno formado por una de las variables, y se calcula la matriz de similitud basada en $\hat{\Delta}^2$:

	X_1	X_2	X_3	X_4	X_5	X_6	X_7
X_1							
X_2	0.02126						
X_3	0.02085	0.02072					
X_4	0.00024	0.00026	0.00021				
X_5	0.01596	0.01671	0.01640	0.05259			
X_6	0.00030	0.00033	0.00031	0.00006	0.01535		
X_7	0.00058	0.00060	0.00062	0.00013	0.01466	0.01133	

Por tanto se unen los clústers $\{X_4\}$ y $\{X_5\}$, ya que muestran la mayor similitud.

- **Paso 2 :** La nueva matriz de similitud viene dada por:

	X_1	X_2	X_3	$\{X_4, X_5\}$	X_6	X_7
X_1						
X_2	0.02126					
X_3	0.02085	0.02072				
$\{X_4, X_5\}$	0.01030	0.01069	0.01049			
X_6	0.00030	0.00033	0.00031	0.01018		
X_7	0.00058	0.00060	0.00062	0.00978	0.01133	

Se unen los clústers $\{X_1\}$ y $\{X_2\}$.

- **Paso 3 :** Se vuelve a calcular la matriz de similitud:

	$\{X_1, X_2\}$	X_3	$\{X_4, X_5\}$	X_6	X_7
$\{X_1, X_2\}$					
X_3	0.02076				
$\{X_4, X_5\}$	0.00607	0.01049			
X_6	0.00382	0.00031	0.01018		
X_7	0.00430	0.00062	0.00978	0.01133	

Se unen los clústers $\{X_1, X_2\}$ y $\{X_3\}$.

- **Paso 4 :** Se vuelve a calcular la matriz de similitud:

	$\{X_1, X_2, X_3\}$	$\{X_4, X_5\}$	X_6	X_7
$\{X_1, X_2, X_3\}$				
$\{X_4, X_5\}$	0.00422			
X_6	0.00456	0.01018		
X_7	0.00496	0.00978	0.01133	

Se unen los clústers $\{X_6\}$ y $\{X_7\}$.

Los tres clústers formados por el algoritmo son $\{X_1, X_2, X_3\}$, $\{X_4, X_5\}$ y $\{X_6, X_7\}$.

Anexos

A | Demostración del teorema de Sklar

Se presenta una prueba alternativa basada en trabajo de los autores F. Durante, J. Fernández-Sánchez y C. Sempi (ver [18]) que además mostrará algunas propiedades del conjunto de las cópulas como subconjuntos de espacio de funciones continuas en $[0, 1]^n$.

| Definición A.1. Se define como \mathcal{C}_n como el conjunto de las cópulas en dimensión n . Se considerará como subconjunto del espacio vectorial $(C([0, 1]^n, \|\cdot\|_\infty))$, que, como es bien sabido, es un espacio métrico completo.

| Teorema A.1. Sea $\{C_k\}_{k \in \mathbb{N}} \subset \mathcal{C}_n$ y $C_0 \in \mathcal{C}_n$ de forma que $C_k \xrightarrow{k} C_0$ puntualmente, entonces $C_k \xrightarrow{k} C_0$ uniformemente.

Demostración. Ver [18]. |

Proposición A.1. El subconjunto \mathcal{C}_n es cerrado en la topología de la convergencia uniforme y, en consecuencia, completo.

Demostración. Considérese una sucesión $\{C_n\}$ que converge uniformemente a C . Hay que ver que $C \in \mathcal{C}_n$. Para comprobar que C es función de distribución es fácil ver que gracias a la convergencia uniforme se verifican las condiciones del *Teorema 1.1*. Se procede a mostrar que las marginales deben ser uniformes, se realizará la prueba para la primera marginal, pues es análogo para el resto.

Sea $\varepsilon > 0$ fijo, por la convergencia uniforme existirá n_0 natural tal que, para todo $n \geq n_0$:

$$\sup_{\underline{u} \in [0,1]^n} |C_n(\underline{u}) - C(\underline{u})| < \varepsilon$$

En particular, para $a_2, \dots, a_n > 0$ arbitrarios se tiene que, para todo $n \geq n_0$ y todo $u_1 \in (0, 1)$:

$$|C_n(u_1, a_2, \dots, a_n) - C(u_1, a_2, \dots, a_n)| < \varepsilon$$

y tomando límites en esta expresión cuando $a_i \rightarrow 1$ para $i = 2, \dots, n$ se obtiene que,

de nuevo para todo $n \geq n_0$, y todo $u_1 \in (0, 1)$:

$$|F_{1,n}(u_1) - F_1(u_1)| = |u_1 - F(u_1)| < \varepsilon,$$

y finalmente de la arbitrariedad de $\varepsilon > 0$ se deduce el resultado. La completitud se desprende de que es un subconjunto cerrado de $(C([0, 1]^n), \|\cdot\|_\infty)$, que es un espacio de Banach. |

| Teorema A.2. *El subconjunto \mathcal{C}_n es compacto en la topología de la convergencia uniforme.*

Demostración. Como \mathcal{C}_n es cerrado bastará ver que es relativamente compacto. Por un lado:

$$\sup \{|C(\underline{u})| : \underline{u} \in [0, 1]^n, C \in \mathcal{C}_n\} \leq 1$$

luego \mathcal{C}_n es uniformemente acotado. Por otro lado, de la *Proposición 1.2* se deduce que dicha familia también es equicontinua, luego aplicando el *teorema de Ascoli-Arzelá* se concluye que \mathcal{C}_n es relativamente compacto. |

Para la demostración se hará uso de una familia regularizante de funciones $\{\rho_\varepsilon\}_{\varepsilon>0}$, esto es, una familia de funciones que verifican, para todo $\varepsilon > 0$, las siguientes propiedades:

1. $\rho_\varepsilon \in C_c^\infty(\mathbb{R}^n)$
2. $\rho_\varepsilon \geq 0$
3. $\text{sop}(\rho_\varepsilon) \subset B(0, \varepsilon)$
4. $\int_{\mathbb{R}^n} \rho_\varepsilon(\underline{x}) d\underline{x} = 1$

Puede comprobarse que la familia dada por $\rho_\varepsilon(\underline{x}) = \frac{1}{\varepsilon^n} \rho(\frac{\underline{x}}{\varepsilon})$ verifica lo anterior, donde

$$\rho(\underline{x}) = \frac{\hat{\rho}(\underline{x})}{\int_{\mathbb{R}^n} \hat{\rho}(\underline{z}) d\underline{z}}$$

siendo

$$\hat{\rho}(\underline{x}) = \begin{cases} e^{\frac{-1}{1-|\underline{x}|^2}}, & \text{si } \underline{x} \in B(0, 1) \\ 0, & \text{si } \underline{x} \notin B(0, 1). \end{cases}$$

Se introduce ahora el concepto de convolución entre funciones, una herramienta muy útil a la hora de obtener funciones con más regularidad a partir de funciones ya dadas.

| Definición A.2. Dadas f, g funciones medibles, entonces, para todo $\underline{x} \in \mathbb{R}^n$ tal que la función $\underline{y} \mapsto f(\underline{x} - \underline{y})g(\underline{y})$ sea integrable, se define:

$$(f * g)(\underline{x}) = \int_{\mathbb{R}^n} f(\underline{x} - \underline{y})g(\underline{y}) d\underline{y}.$$

A esta nueva función la llamamos la convolución en f con g .

Observación A.1. Si la convolución de f con g está bien definida en el punto \underline{x} , entonces mediante el cambio de variables $\underline{u} = \underline{x} - \underline{y}$ se tiene que $g * f$ también está bien definida en el punto \underline{x} y, de hecho, $(f * g)(\underline{x}) = (g * f)(\underline{x})$, lo cual muestra la conmutatividad del producto de convolución.

Proposición A.2. Sea H función de distribución n -dimensional. Para todo $k \in \mathbb{N}$ definimos

$$H_k(\underline{x}) = \int_{\mathbb{R}^n} H(\underline{x} - \underline{y})\rho_{1/k}(\underline{y}) d\underline{y}.$$

Entonces se verifica que H_k es función de distribución n -dimensional continua para todo $k \in \mathbb{N}$. En consecuencia, cada función de distribución marginal $F_{k,i}$ de H_n es continua.

Demostración. El hecho de que $0 \leq H \leq 1$ indica que H_k está bien definida y toma valores entre 0 y 1 para todo $k \in \mathbb{N}$. Para ver que es función de distribución bastará comprobar las condiciones del *Teorema 1.1*. Si se fija $\varepsilon > 0$ arbitrario, por ser H función de distribución deberá existir $M > 0$ tal que, para todo $\underline{X} \in \mathbb{R}^n$ verificando $x_i > M$ para $i = 1, \dots, n$, se cumple $H(\underline{x}) > 1 - \varepsilon$. Se sigue de aquí que, para todo $\underline{x} \in \mathbb{R}^n$ verificando $x_i > M + 1/k$ para $i = 1, \dots, n$ y dado \underline{y} en $B(0, 1/k)$ entonces $x_i - y_i > M$, con lo cual

$$H_k(\underline{x}) = \int_{B(0, 1/k)} H(\underline{x} - \underline{y})\rho_{1/k}(\underline{y}) d\underline{y} > (1 - \varepsilon) \int_{B(0, 1/k)} \rho_{1/k}(\underline{y}) d\underline{y} = 1 - \varepsilon$$

de donde se deduce que $H_k(\underline{x}) \rightarrow 1$ si $x_i \rightarrow +\infty$ para $i = 1, \dots, n$. De forma similar se prueba $H_k(\underline{x}) \rightarrow 0$ si $x_i \rightarrow -\infty$ para algún $i \in \{1, \dots, n\}$. Para ver que se verifica la segunda propiedad del *Teorema 1.1*, basta considerar $[\underline{a}, \underline{b}]$, y notar que :

$$\begin{aligned} V_{H_n}([\underline{a}, \underline{b}]) &= \sum_{\underline{v}} \text{sign}(\underline{v}) H_n(\underline{v}) = \int_{\mathbb{R}^n} \sum_{\underline{v}} \text{sign}(\underline{v}) H(\underline{v} - \underline{y})\rho_{1/k}(\underline{y}) d\underline{y} \\ &= \int_{\mathbb{R}^n} V_H[\underline{a} - \underline{y}, \underline{b} - \underline{y}]\rho_{1/k}(\underline{y}) d\underline{y} \geq 0, \end{aligned}$$

donde el último término es no negativo por ser la integral de una función no negativa, ya que H es una función de distribución y verifica la desigualdad en los puntos $\underline{a} - \underline{y}$

y $\underline{b} - \underline{y}$. Resta ver que H_k es continua, como $\rho_{1/k}$ es continua en $\overline{B(0, 1)}$ compacto, es uniformemente continua aquí, por tanto dado $\varepsilon > 0$ existe $\delta > 0$ tal que si $|\underline{x} - \underline{y}| < \delta$ se cumple $|\rho_{1/k}(\underline{x}) - \rho_{1/k}(\underline{y})| < \varepsilon$. Así, denotando como m_n a la medida de Lebesgue n -dimensional, y usando la conmutatividad de la convolución, así como el hecho de que $|H| \leq 1$:

$$|H_k(\underline{x}) - H_k(\underline{u})| \leq \int_{B(0, 1/k)} |H(\underline{y})| |\rho_{1/k}(\underline{x} - \underline{y}) - (\underline{x} - \underline{y})| d\underline{y} \leq 2\varepsilon m_n(B(0, 1/k))$$

de donde se deduce la continuidad. |

Proposición A.3. Si H es continua en el punto $\underline{x} \in \mathbb{R}^n$, entonces

$$\lim_{k \rightarrow +\infty} H_k(\underline{x}) = H(\underline{x}).$$

En consecuencia, si H tiene marginales F_1, \dots, F_n y F_i es continua en el punto $x_i \in \mathbb{R}$, entonces

$$\lim_{k \rightarrow +\infty} F_{k,i}(x_i) = F_i(x_i).$$

Demostración. Si H es continua en el punto $\underline{x} \in \mathbb{R}^n$, entonces dado $\varepsilon > 0$ existe $\delta > 0$ tal que para todo $\underline{y} \in B(0, \delta)$ se verifica $|H(\underline{x}) - H(\underline{x} - \underline{y})| < \varepsilon$. Por tanto, para todo $k > 1/\delta$:

$$\begin{aligned} |H_k(\underline{x}) - H(\underline{x})| &= \left| \int_{B(0, 1/k)} H(\underline{x} - \underline{y}) \rho_{1/k}(\underline{y}) d\underline{y} - H(\underline{x}) \int_{B(0, 1/k)} \rho_{1/k}(\underline{y}) d\underline{y} \right| \\ &= \left| \int_{B(0, 1/k)} (H(\underline{x} - \underline{y}) - H(\underline{x})) \rho_{1/k}(\underline{y}) d\underline{y} \right| \\ &\leq \int_{B(0, 1/k)} |H(\underline{x} - \underline{y}) - H(\underline{x})| \rho_{1/k}(\underline{y}) d\underline{y} \\ &< \varepsilon \int_{B(0, 1/k)} \rho_{1/k}(\underline{y}) d\underline{y} = \varepsilon, \end{aligned}$$

de donde se deduce el resultado. |

Ya con todas las herramientas listas se procede a la demostración del *Teorema de Sklar*:

Teorema A.3 (Sklar). Sea X un vector aleatorio con función de distribución H y funciones de distribución marginales univariantes F_1, \dots, F_n . Entonces existe una cópula C asociada con X , es decir, tal que se verifica que

$$H(\underline{x}) = C(F_1(x_1), \dots, F_n(x_n)).$$

para todo $\underline{x} \in \mathbb{R}^n$.

Demostración. Dada la función de distribución H se construye la sucesión de funciones de distribución continuas H_k , cuyas marginales $F_{k,1}, \dots, F_{k,n}$ también son continuas. Aplicando a cada H_k la *Proposición 1.4* se obtiene que existe una sucesión de cópulas $\{C_k\}$ tal que, para todo $k \in \mathbb{N}$:

$$H_k(\underline{x}) = C_k(F_{k,1}(x_1), \dots, F_{k,n}(x_n)).$$

Por la compacidad de \mathcal{C}_n , existe una subsucesión $\{C_{k(m)}\}$ de $\{C_k\}$ que converge uniformemente a una cópula C . Sea entonces $\underline{x} = (x_1, \dots, x_n) \in \mathbb{R}^n$, se distinguen aquí dos casos:

- H es continua en \underline{x} :

Aplicando la *Proposición A.3*, se tiene que $\{H_k(\underline{x})\}$ converge a $H(\underline{x})$, y por tanto también lo hace $\{H_{k(m)}(\underline{x})\}$. Este hecho también se aplica a las funciones de distribución marginales, con lo que la sucesión

$$\{C_{k(m)}(F_{k(m),1}(x_1), \dots, F_{k(m),n}(x_n))\}$$

converge a $C(F_1(x_1), \dots, F_n(x_n))$. De esta forma:

$$\begin{aligned} H(\underline{x}) &= \lim_m H_{k(m)}(\underline{x}) = \lim_m \{C_{k(m)}(F_{k(m),1}(x_1), \dots, F_{k(m),n}(x_n))\} \\ &= C(F_1(x_1), \dots, F_n(x_n)), \end{aligned}$$

con lo que se verifica el teorema.

- H no es continua en \underline{x} :

Se escoge en este caso una sucesión de puntos de continuidad de H , $\{\underline{x}_j\}$, verificando $x_{j,i} > x_i$ para $i = 1, \dots, n$ y todo $j \in \mathbb{N}$, y tales que $\{H(\underline{x}_j)\}$ converge a $H(\underline{x})$ y $\{F_i(x_{j,i})\}$ converge a $F_i(x_i)$ para $i = 1, \dots, n$ (nótese que esto es posible por ser las funciones de distribución continuas a la derecha). Por tanto, dado $\varepsilon > 0$, existirá $j_0 \in \mathbb{N}$ tal que, para todo $j \geq j_0$, se tiene

$$|H(\underline{x}_j) - H(\underline{x})| < \varepsilon \text{ y } |F_i(x_{j,i}) - F_i(x_i)| < \varepsilon/(2n), \text{ para } i = 1, \dots, n.$$

Por otro lado, como dado \underline{x}_j es un punto de continuidad, gracias a la *Proposición A.3*, para un $m \in \mathbb{N}$ suficientemente grande, se tendrá que

$$|H_{k(m)}(\underline{x}_j) - H(\underline{x}_j)| < \varepsilon.$$

Por ello

$$|H_{k(m)}(\underline{x}_j) - H(\underline{x})| \leq |H_{k(m)}(\underline{x}_j) - H(\underline{x}_{j_0})| + |H(\underline{x}_j) - H(\underline{x})| < 2\varepsilon,$$

de donde se deduce que $\{H_{k(m)}(\underline{x}_j)\}_{m,j}$ converge a $H(\underline{x})$ cuando $m, j \rightarrow +\infty$. Además la sucesión

$$\{C_{k(m)}(F_1(x_{j,1}), \dots, F_n(x_{j,n}))\}_{m,j}$$

converge a $C(F_1(x_1), \dots, F_n(x_n))$ cuando $m, j \rightarrow +\infty$. En efecto, para j, m suficientemente grandes, si aplicamos la desigualdad triangular:

$$\begin{aligned} & |C_{k(m)}(F_1(x_{j,1}), \dots, F_n(x_{j,n})) - C(F_1(x_1), \dots, F_n(x_n))| \leq \\ & |C_{k(m)}(F_1(x_{j,1}), \dots, F_n(x_{j,n})) - C_{k(m)}(F_1(x_1), \dots, F_n(x_n))| \\ & + |C(F_1(x_1), \dots, F_n(x_n)) - C_{k(m)}(F_1(x_1), \dots, F_n(x_n))| = (*) \end{aligned}$$

acotando el primer sumando por la *Proposición 1.2* y el segundo por la convergencia uniforme de $\{C_{k(m)}\}$:

$$(*) \leq \sum_{i=1}^n |F_i(x_{j,i}) - F_i(x_i)| + \frac{\varepsilon}{2} \leq n \frac{\varepsilon}{2n} + \frac{\varepsilon}{2} = \varepsilon$$

que era lo que se quería ver. Finalmente, usando el resultado ya probado para puntos de continuidad:

$$H(\underline{x}) = \lim H_{k(m)}(\underline{x}_j) = \lim C_{k(m)}(F_1(x_{j,1}), \dots, F_n(x_{j,n})) = C(F_1(x_1), \dots, F_n(x_n))$$

concluyendo así la prueba. |

B | La cópula empírica

La siguiente exposición se basa en la sección 4.2 de [17]. Sea X un vector aleatorio d -dimensional, con funciones de distribución marginales F_1, \dots, F_d y cópula C , del que se posee una muestra aleatoria independiente X_1, \dots, X_n . Para cada una de estas funciones de distribución marginales, un estimador no paramétrico clásico es el dado por la función de distribución empírica, con expresión

$$F_{n,j}(x) = \frac{1}{n+1} \sum_{i=1}^n \mathbb{I}(X_{ij} \leq x), \quad \forall x \in \mathbb{R},$$

donde \mathbb{I} representa la función indicatriz y $j \in \{1, \dots, d\}$.

Haciendo uso dichos estimadores, podemos definir la muestra asociada de pseudo-observaciones de C como

$$U_{i,n} = (F_{n,1}(X_{i1}), \dots, F_{n,d}(X_{id})), \quad i \in \{1, \dots, n\}.$$

Nótese que $U_{1,n}, \dots, U_{n,n}$ no son observaciones reales de C , de hecho no son observaciones independientes. La razón de dividir por $n+1$ en la definición de la función de distribución empírica es que, debido a que estas pseudo-observaciones serán usadas para estimar la cópula de X , es necesario que pertenezcan al interior de hipercubo $[0, 1]^d$. Aunque de esta forma dichas funciones de distribución empíricas no pueden tomar el valor 1, esto no influye para grandes volúmenes de datos, pues $\frac{n}{n+1} \approx 1$ para valores elevados de n .

La cópula empírica C_n consiste en un estimador no paramétrico de C , dado por

$$C_n(\underline{u}) = \frac{1}{n} \sum_{i=1}^n \mathbb{I}(U_{i,n} \leq \underline{u}) = \frac{1}{n} \sum_{i=1}^n \prod_{j=1}^d \mathbb{I}(U_{ij,n} \leq \underline{u}), \quad \underline{u} \in [0, 1]^d.$$

Deheuvels probó en 1979 que C_n es un estimador consistente de C en [19], sin embargo a veces muestra un gran sesgo cuando el tamaño de muestra es pequeño. Para

solucionar este problema, Sancetta y Satchell introdujeron una nueva clase de estimadores de C más suaves, las cópulas empíricas de Bernstein (ver [20]). Dentro de esta clase destaca la *Cópula empírica beta*, que fue introducida por Segers en [21], y cuya expresión viene dada por

$$C_n^\beta(\underline{u}) = \frac{1}{n} \sum_{i=1}^n \prod_{j=1}^d F_{n,R_{ij}}(u_j), \quad \underline{u} \in [0, 1]^d,$$

donde, para cada $r \in \{1, \dots, n\}$, $F_{n,r}$ denota la función de distribución de la variable aleatoria Beta con parámetros r y $n + 1 - r$, y donde, para cada $j \in \{1, \dots, d\}$, R_{ij} representa el rango de X_{ij} en X_{1j}, \dots, X_{nj} , dado por

$$R_{ij} = \sum_{k=1}^n \mathbb{I}(X_{kj} < X_{ij}).$$

Otro estimador no paramétrico suave de C es la *cópula empírica "checkerboard"*, definida como

$$C_n^\#(\underline{u}) = \frac{1}{n} \sum_{i=1}^n \prod_{j=1}^d \min \{ \max \{ nu_j - R_{ij} + 1, 0 \}, 1 \}, \quad \underline{u} \in [0, 1]^d.$$

Este estimador es una gran herramienta, sobre todo en el contexto de datos discontinuos, para más detalles ver [22].

Bibliografía

- [1] B. Schweizer, A. Sklar: *Probabilistic Metric Spaces*.
Dover Publications Inc, New York, 2005.
- [2] Krishna B. Athreya, Soumendra N. Lahiri: *Measure Theory and Probability Theory*.
Springer, New York, 2006.
- [3] Harry Joe: *Multivariate Models and Multivariate Dependence Concepts*.
Chapman and Hall CRC, London, 1997.
- [4] Roger B. Nelsen: *An Introduction to Copulas*.
Springer, New York, 1999.
- [5] Victoriano J. García García, Héctor M. Ramos Romero, Miguel Ángel Sordo Díaz:
193 problemas resueltos de cálculo de probabilidades.
UCA, Cádiz, 2008.
- [6] H.L Royden: *Real Analysis, Second Edition*.
The Macmillan Company, London, 1968.
- [7] R.T. Seeley: *Fubini implies Leibniz implies $F_{yx} = F_{xy}$* .
The American Mathematical Monthly 68, pages 56-57. 1961.
- [8] Fabrizio Durante, Carlo Sempi, Piotr Jaworski, Fabrizio Durante, Wolfgang Karl Härdle, Tomasz Rychlik: *Copula Theory and Its Applications*.
Springer, Heidelberg, 2009.
- [9] C. Spearman: *The proof and measurement of association between two things*.
American Journal of Psychology 15, pages 72-101. 1904.

- [10] Harry Joe: *Relative Entropy Measures of Multivariate Dependence*.
Journal of the American Statistical Association Vol. 84, No. 405, pages 157-164.
1989.
- [11] W. Hoeffding: *Massstabinvariante Korrelationstheorie*.
Schriften des Mathematischen Seminars und des Instituts für Angewandte Mathematik der Universität Berlin 5(3), pages 181–233. 1940.
- [12] B. Schweizer, E.F. Wolff: *On nonparametric measures of dependence for random variables*.
The Annals of Statistics Vol.9 N°4, pages 879–885. 1981.
- [13] J. Pickands: *Multivariate extreme value distributions*.
Proceedings of the 43rd session of the International Statistical Institute Vol. 2, pages 859–878, 894–902. 1981.
- [14] G. Kimeldorf, A. R. Sampson: *Uniform representations of bivariate distributions*.
Communications in Statistics, pages 617-627. 1975.
- [15] Ioannis Kosmidis, Dimitris Karlis: *Model-based clustering using copulas with applications*.
Statistics and Computing volume 26, pages 1079–1099. 2015.
- [16] Sylvia Frühwirth-Schnatter: *Finite Mixture and Markov Switching Models*.
Springer, New York, 2015.
- [17] Marius Hofert, Ivan Kojadinovic, Martin Mächler, Jun Yan: *Elements of Copula Modeling with R*.
Springer, Zürich, 2018.
- [18] F. Durante, J. Fernández-Sánchez, C. Sempi: *Sklar's Theorem obtained via regularization techniques*.
Nonlinear Anal-Theor 75, pages 769-774. 2012.
- [19] P. Deheuvels: *La fonction de dépendance empirique et ses propriétés: un test non paramétrique d'indépendance*.
Académie Royale de Belgique Bulletin de la Classe des Sciences 5th Series 65, pages 274–292. 1979.

- [20] A. Sancetta, S. Satchell: *The Bernstein copula and its applications to modeling and approximations of multivariate distributions*.
Econometric Theory 20, pages 535–562. 2004.
- [21] J. Segers, M. Sibuya, H. Tsukahara: *The empirical beta copula*.
Journal of Multivariate Analysis 155, pages 35–51. 2017.
- [22] H. Carley, M. D. Taylor: *A new proof of Sklar's theorem*.
in C. M. Cuadras, J. Fortiana, J. A. Rodríguez-Lallena: *Distributions with given marginals and statistical modelling*.
Springer, Dordrecht, 2002.
- [23] C. Genest, L. P. Rivest: *Statistical inference procedures for bivariate Archimedean copulas*.
Journal of the American Statistical Association 88, pages 1034–1043. 1993.