

---

---

# UNA INTRODUCCIÓN A LA TEORÍA DE CÓPULAS CON APLICACIONES A LA HIDROLOGÍA Y LAS FINANZAS

---

---

TRABAJO FIN DE GRADO

Autor:

Emilio Gil Martínez

Tutor:

Enrique de Amo Artero

GRADO EN MATEMÁTICAS



JULIO, 2016  
Universidad de Almería



# Índice general

<b>1</b>	<b>Cópulas</b>	<b>3</b>
1.1.	<b>Nociones sobre cópulas</b>	<b>3</b>
	Definición de cópula, 3.— Primeras propiedades, 3.	
1.2.	<b>Teorema de Sklar</b>	<b>5</b>
	El Teorema de Sklar, 5.	
1.3.	<b>Medidas de asociación</b>	<b>9</b>
	Concordancia, 10.— Tau de Kendall, 10.— Rho de Spearman, 13.— Dependencia de colas, 15.	
1.4.	<b>Familias de cópulas</b>	<b>16</b>
	La cópula empírica, 16.— Cópulas paramétricas, 17.	
<b>2</b>	<b>Aplicación en finanzas</b>	<b>19</b>
	Introducción, 19.— Estimación de las marginales, 19.— Estudio de la dependencia, 20.— Selección de la cópula: estimación, 21.— Selección de la cópula: test de ajuste, 21.— Haciendo predicciones, 22.— Resultados simulados, 23.— Otras consideraciones, 23.	
<b>3</b>	<b>Aplicación en hidrología</b>	<b>25</b>
	Introducción, 25.— Estimación de las marginales, 25.— Estudio de la dependencia, 25.— Selección de la cópula: estimación y test de ajuste, 25.— Resultados simulados, 28.	
<b>4</b>	<b>Conclusiones</b>	<b>33</b>
	<b>Bibliografía</b>	<b>35</b>



# Resumen

Nuestra meta en este trabajo es entender cómo esa herramienta tan sencilla que son las cópulas ofrece soluciones bastante efectivas en multitud de campos. Para ello, vamos a estudiar el concepto de cópula de forma similar a como propone Nelsen [7] en el Capítulo 1, haciendo especial hincapié en el teorema de Sklar. También trabajaremos temas relacionados con las cópulas como son las medidas de asociación o las familias de cópulas, de cara a comprender los métodos que se usarán a continuación. Un primer uso de toda esta teoría lo encontramos en el artículo de Grégoire [4] que desarrollamos en el Capítulo 2, donde usamos algunos de los métodos más simples en el modelado con cópulas para estudiar el mercado energético. Tras esto, en el Capítulo 3 veremos como se pueden abordar problemas en hidrología. En particular, Requena se ha enfrentado a la predicción de inundaciones en su artículo [9] ampliando así el abanico metodológico. Esperamos dar así una visión amplia de los importantes resultados que se pueden conseguir con las cópulas de forma sencilla y eficaz. Debido al carácter de las cópulas, se han usado nociones de análisis real en el marco teórico, mientras que, en la parte más práctica, se han utilizado conceptos tanto probabilísticos como estadísticos.



# *Abstract*

Our goal in this project is to understand how such an easy tool as copulas can offer very effective solutions in many fields. For this, we will study the concept of copula in a similar way as Nelsen [7] suggests in Chapter 1, with a special effort in Sklar's Theorem. We will also work with topics related to copulas as measures of association or families of copulas in order to understand the methods that will be used next. A first used of all this theory is found in Grégoire's article [4] that we relate in Chapter 2, where some of the simplest methods in modelling with copulas are used in the energy market. After this, in Chapter 3, we will see how to treat with hidrology problems. In particular, Requena faces dam prediction in her article [9] increasing our methods. We hope that this gives a wide view of how important results can be obtained with copulas in a simple and eficiente way. Because of copulas nature, real analysis notions has been used in the theory, while probabilitics and statistics concepts have been used in the practical part.





## *Antecedentes bibliográficos*

El problema de expresar las relaciones de dependencia de las variables aleatorias que definen un vector aleatorio no es nada nuevo. En 1959, Abe Sklar propuso una solución a dicho problema acuñando el término de “cópulas” y resolviéndolo eficazmente. Estas funciones llamadas cópulas permiten expresar dicha dependencia siendo a su vez ciertamente manejables, ya que, son funciones continuas del cuadrado unidad en el intervalo unidad.

Es debido a esto último que la Teoría de Cópulas se ha revelado como una herramienta de extraordinaria eficacia a la hora de estudiar la dependencia o independencia en modelos estocásticos multidimensionales. Su aplicación abarca una gran cantidad de áreas entre las que destacan la hidrología o las finanzas.



# Cóputas

Dedicamos este capítulo al concepto central de este trabajo, las cóputas. A su vez hacemos especial hincapié en el teorema de Sklar, en el que se establece la relación entre una cóputa y la función de distribución conjunta de un vector aleatorio a través de sus marginales. También veremos herramientas como las medidas de asociación relacionándolas con las cóputas y una serie de cóputas de gran utilidad.

## 1.1 Nociones sobre cóputas

### Definición de cóputa

Para empezar nuestro trabajo lo primero que vamos a hacer es definir qué es una cóputa para poder extraer sus primeras propiedades y fijar la notación usada.

**Definición 1.1.** Una cóputa  $n$ -dimensional  $C$  es una función definida en  $[0, 1]^n$  cuya imagen es  $[0, 1]$  con las siguientes propiedades:

$$\forall 1 \leq k \leq n \quad C(u_1, \dots, u_{k-1}, 0, u_{k+1}, \dots, u_n) = 0 \quad (1.1)$$

$$\forall 1 \leq k \leq n \quad u_k \in [0, 1] \quad C(1, \dots, 1, u_k, 1, \dots, 1) = u_k \quad (1.2)$$

$$\forall u_1, \dots, u_n, v_1, \dots, v_n \in [0, 1] : u_i \leq v_i, i = 1, \dots, n$$

$$\sum_{i_1=1}^2 \sum_{i_n=1}^2 (-1)^{i_1+\dots+i_n} C(u_{1,i_1}, \dots, u_{n,i_n}) \geq 0 \quad (1.3)$$

donde  $u_{j,1} = u_j$  y  $u_{j,2} = v_j$ .

Aunque la definición, como vemos, está dada para dimensión  $n \geq 2$ , casi siempre trabajaremos con dimensión dos. Por tanto, de ahora en adelante cuando hablemos de cóputa nos referiremos a una cóputa bidimensional. Introducimos ahora un concepto muy similar relajando el dominio de  $C$ , llegamos así a la noción de subcóputa.

**Definición 1.2.** Una subcóputa  $C'$  es la restricción de una cóputa al conjunto  $S_1 \times S_2$  con  $S_1, S_2 \subset [0, 1]$  y con  $0, 1 \in S_1, S_2$

Este concepto es más teórico que práctico ya que lo usaremos para probar más adelante el Teorema de Sklar. Añadimos ahora un concepto asociado al de cóputa bastante usado en el plano de las aplicaciones. Este es el de la densidad de una cóputa  $C$ .

**Definición 1.3.** Sea  $C$  una cóputa, decimos que  $c$  es su función de densidad de cóputa si:

$$c(u, v) = \frac{\partial^2 C}{\partial u \partial v}(u, v). \quad (1.4)$$

### Primeras propiedades

Antes de seguir profundizando en el concepto de cóputa es necesario saber que este concepto es aplicable, es decir, existen funciones que respondan a la definición que hemos dado de cóputa.

**Proposición 1.1.** Las funciones  $W(u, v) = \max(u+v-1, 0)$ ,  $M(u, v) = \min(u, v)$  y  $\Pi(u, v) = uv$  con dominio  $[0, 1] \times [0, 1]$  son cópulas.

Demostración:

Demostraremos solo que la función  $W$  es una cópula, las otras dos se hacen de forma similar. Es claro que la cópula es simétrica, luego para probar las dos primeras propiedades solo las demostraremos para una de las variables.

$$W(u, 0) = \max(u-1, 0) = 0 \text{ ya que } u \leq 1$$

$$W(u, 1) = \max(u, 0) = u \text{ ya que } u \geq 0$$

$$\forall u_1, u_2, v_1, v_2 \in [0, 1]: u_1 \leq u_2, v_1 \leq v_2$$

$$W(u_2, v_2) - W(u_1, v_2) - W(u_2, v_1) + W(u_1, v_1)$$

$$= \max(u_2 + v_2 - 1, 0) - \max(u_1 + v_2 - 1, 0) - \max(u_2 + v_1 - 1, 0) + \max(u_1 + v_1 - 1, 0).$$

Distinguimos 6 posibilidades (que el primer máximo sea 0 fuerza todos los demás a serlo por la hipótesis de  $u_1 \leq u_2, v_1 \leq v_2$  y de igual forma con los dos intermedios):

$$0 - 0 - 0 + 0 = 0 \geq 0$$

$$(u_2 + v_2 - 1) - 0 - 0 + 0 = u_2 + v_2 - 1 \geq 0$$

$$(u_2 + v_2 - 1) - (u_1 + v_2 - 1) - 0 + 0 = u_2 - u_1 \geq 0$$

$$(u_2 + v_2 - 1) - 0 - (u_2 + v_1 - 1) + 0 = v_2 - v_1 \geq 0$$

$$(u_2 + v_2 - 1) - (u_1 + v_2 - 1) - (u_2 + v_1 - 1) + 0 = 1 - u_1 - v_1 \geq 0$$

$$(u_2 + v_2 - 1) - (u_1 + v_2 - 1) - (u_2 + v_1 - 1) + (u_1 + v_1 - 1) = 0 \geq 0.$$

■

Estas tres cópulas son de vital importancia dentro del marco teórico, vemos entonces la propiedad principal de las cópulas  $W$  y  $M$  antes de darles nombres.

**Proposición 1.2** (Cotas de Frechet-Hoeffding). Sea  $C$  una cópula se tiene que dicha cópula verifica la relación:

$$W(u, v) \leq C(u, v) \leq M(u, v), \forall u, v \in [0, 1]$$

Demostración:

Sea  $u, v \in [0, 1]$  se tiene que  $C(u, v) \leq C(u, 1) = u$  y  $C(u, v) \leq C(1, v) = v$ , luego  $C(u, v) \leq \min(u, v) = M(u, v)$ . Por otra parte, para  $u, v \in [0, 1]$  se tiene que  $C(1, 1) - C(u, 1) - C(1, v) + C(u, v) = 1 - u - v + C(u, v) \geq 0$  lo que, despejando, nos da que  $C(u, v) \geq u + v - 1$  y como además tenemos que  $C(u, v) \geq 0$  tenemos que  $C(u, v) \geq \max(u + v - 1, 0) = W(u, v)$ , lo que finaliza la demostración.

■

Debido a esto, a las cópulas  $M$  y  $W$  se les da el nombre de cópulas maximal y minimal respectivamente.

## 1.2 Teorema de Sklar

### El Teorema de Sklar

La principal razón para utilizar las cópulas en el mundo de la estadística es como veremos a continuación darle un sentido numérico a la relación que pueda existir entre dos variables aleatorias.

Para poder entender de qué habla el Teorema de Sklar primero tenemos que definir qué es la función de distribución asociada a una variable aleatoria y lo que es la función de distribución asociada a un vector aleatorio.

**Definición 1.4.** Una función de distribución  $F$  es aquella cuyo dominio es  $\overline{\mathbb{R}}$  y que cumple que:

$$F \text{ es no decreciente} \quad (1.5)$$

$$F(-\infty) = 0 \text{ y } F(+\infty) = 1. \quad (1.6)$$

Esta sería la definición general. Sin embargo, en el caso de que tengamos un vector aleatorio, varias funciones de distribución son identificables asociadas a dicho vector. Como siempre, asumimos que estamos en el caso bidimensional:

**Definición 1.5.** Una función de distribución bidimensional  $H$  es aquella cuyo dominio es  $\overline{\mathbb{R}^2}$  y que cumple que:

$$H \text{ es monótona, es decir, sus volúmenes son no negativos} \quad (1.7)$$

$$H(x, -\infty) = H(-\infty, y) = 0 \text{ y } H(+\infty, +\infty) = 1. \quad (1.8)$$

Además, las funciones dadas por  $F(x) = H(x, +\infty)$  y  $G(y) = H(+\infty, y)$  son también funciones de distribución llamadas marginales de  $H$ .

La relación de estas funciones con las variables aleatorias es la siguiente:

$$H(x, y) = \mathbb{P}(X \leq x, Y \leq y)$$

$$F(x) = \mathbb{P}(X \leq x)$$

$$G(y) = \mathbb{P}(Y \leq y).$$

Con el fin de más adelante ver la versión de densidades del Teorema de Sklar definimos las funciones de densidad.

**Definición 1.6.** Sean  $F$ ,  $G$  y  $H$  como en la definición anterior. Llamamos funciones de densidad asociadas a  $f$ ,  $g$  y  $h$  dadas por:

$$f(x) = \frac{\partial}{\partial x} F(x)$$

$$g(y) = \frac{\partial}{\partial y} G(y)$$

$$h(x, y) = \frac{\partial^2}{\partial x \partial y} H(x, y).$$

Pasamos ahora a dar unos lemas previos para demostrar el Teorema de Sklar.

**Lema 1.1.** *Sea  $H$  una función de distribución conjunta y  $F$  y  $G$  sus marginales, entonces existe una única subcópula  $C'$  tal que:*

$$\text{Dom } C' = \text{Ran } F \times \text{Ran } G$$

$$\forall x, y \in \overline{\mathbb{R}}, H(x, y) = C'(F(x), G(y)).$$

Demostración:

Por la definición dada de  $H(x, y)$  afirmamos que para todo par de puntos  $(x_1, y_1)$  y  $(x_2, y_2)$  se tiene

$$|H(x_2, y_2) - H(x_1, y_1)| \leq |H(x_2, y_2) - H(x_1, y_2)| + |H(x_1, y_2) - H(x_1, y_1)|$$

por la desigualdad triangular. Por otra parte usando que el volumen de  $H$  es positivo podemos decir que si  $x_2 \geq x_1$  se tiene que:

$$H(x_2, \infty) - H(x_1, \infty) - H(x_2, y_2) + H(x_1, y_2) \geq 0$$

$$\Rightarrow H(x_2, \infty) - H(x_1, \infty) \geq H(x_2, y_2) - H(x_1, y_2)$$

$$\Rightarrow F(x_2) - F(x_1) \geq H(x_2, y_2) - H(x_1, y_2).$$

De igual forma, podríamos haber supuesto  $x_1 \geq x_2$  y hacer el mismo procedimiento para  $y_1$  e  $y_2$  para sustituir en la ecuación anterior obteniendo:

$$|H(x_2, y_2) - H(x_1, y_1)| \leq |F(x_2) - F(x_1)| + |G(y_2) - G(y_1)|.$$

Además, si  $F(x_1) = F(x_2)$  y  $G(y_1) = G(y_2)$ , se tiene que  $H(x_1, y_1) = H(x_2, y_2)$ , por lo que la asignación

$$(F(x), G(y)) \rightarrow H(x, y)$$

es unívoca. Es decir, existe una función  $C'$  tal que  $C'(F(x), G(y)) = H(x, y)$ . Por definición de  $C'$ , su volumen es positivo, luego cumple 1.3, y teniendo en cuenta

$$C'(F(x), 1) = C'(F(x), G(+\infty)) = H(x, +\infty) = F(x)$$

$$C'(F(x), 0) = C'(F(x), G(-\infty)) = H(x, -\infty) = 0$$

queda demostrado que  $C'$  es una subcópula con dominio  $\text{Ran } F \times \text{Ran } G$ . ■

El siguiente lema nos permite extender una subcópula a una cópula, por lo que se puede usar al lema anterior.

**Lema 1.2.** *Sea  $C'$  una subcópula existe una cópula  $C$  tal que  $C(u, v) = C'(u, v)$  con  $(u, v) \in \text{Dom } C'$ .*

Demostración:

Sea  $C'$  una subcópula con dominio  $S_1 \times S_2$ . De la misma forma que en el lema anterior se puede probar que

$$|C'(x_2, y_2) - C'(x_1, y_1)| \leq |x_2 - x_1| + |y_2 - y_1|$$

ya que  $C'$  también es monótona. Podemos entonces extender por continuidad  $C'$  a una subcópula  $C''$  con dominio  $\overline{S_1} \times \overline{S_2}$ , donde  $\overline{S_1}$  y  $\overline{S_2}$  son los cierres de  $S_1$  y  $S_2$  respectivamente. Es claro que  $C''$  es también una subcópula. Extendemos ahora  $C''$  a una cópula  $C$  de la siguiente forma. Sea  $(a, b)$  perteneciente a  $[0, 1] \times [0, 1]$ , sean  $a_1$  y  $a_2$  el mayor y el menor de los puntos de  $\overline{S_1}$  que cumplen  $a_1 \leq a \leq a_2$  y sean  $b_1$  y  $b_2$  los elementos simétricos de la otra variable. Definimos ahora las funciones interpolatorias  $\lambda$  y  $\mu$  de  $[0, 1]$  en sí mismo:

$$\lambda(a) = \begin{cases} (a - a_1)/(a_2 - a_1), & \text{para } a_1 < a_2 \\ 1, & \text{para } a_1 = a_2 \end{cases}$$

$$\mu(b) = \begin{cases} (b - b_1)/(b_2 - b_1), & \text{para } b_1 < b_2 \\ 1, & \text{para } b_1 = b_2 \end{cases}.$$

Podemos ahora definir la cópula  $C$  como sigue:

$$C(a, b) = (1 - \lambda(a))(1 - \mu(b))C''(a_1, b_1) + (1 - \lambda(a))\mu(b)C''(a_1, b_2) \\ + \lambda(a)(1 - \mu(b))C''(a_2, b_1) + \lambda(a)\mu(b)C''(a_2, b_2).$$

Con  $a_1, a_2, b_1$  y  $b_2$  los definidos anteriormente. Es claro que el dominio de  $C$  es  $[0, 1] \times [0, 1]$  y que además verifica 1.1 y 1.3. Falta probar la monotonía de  $C$ , es decir, comprobar 1.3. Sean  $a, b, c$  y  $d$  pertenecientes a  $[0, 1]$  con  $c \geq a$  y  $d \geq b$ . Existen entonces 4 casos dependiendo de si hay algún punto de  $\overline{S_1}$  estrictamente entre  $a$  y  $c$  o si hay algún punto de  $\overline{S_2}$  entre  $b$  y  $d$ . Supongamos el caso en el que no hay puntos del dominio de la subcópula intermedios, entonces tendríamos que  $a_1 = c_1, a_2 = c_2, b_1 = d_1$  y  $b_2 = d_2$ , teniendo entonces que:

$$C(c, d) - C(a, d) - C(c, b) + C(a, b) \\ = (\lambda(c) - \lambda(a))(\mu(d) - \mu(b))(C''(a_2, b_2) - C''(a_1, b_2) - C''(a_2, b_1) + C''(a_1, b_1)).$$

Donde sabemos que los tres factores son positivos ya que  $\lambda$  y  $\mu$  son no decrecientes y  $C''$  es monótona por ser subcópula. Vemos ahora el caso en el que hay puntos intermedios en las dos variables (el caso en el que solo hay puntos intermedios en una de las variables se haría de forma similar a este y al anterior). Tenemos entonces  $a_1 < a < a_2 < c_1 < c < c_2$  y  $b_1 < b < b_2 < d_1 < d < d_2$ , lo que nos da tras reordenar convenientemente:

$$C(c, d) - C(a, d) - C(c, b) + C(a, b) \\ = (1 - \lambda(a))(1 - \mu(b))(C''(a_2, b_2) - C''(a_1, b_2) - C''(a_2, b_1) + C''(a_1, b_1)) \\ + (1 - \lambda(a))\mu(d)(C''(a_2, d_2) - C''(a_1, d_2) - C''(a_2, d_1) + C''(a_1, d_1)) \\ + \lambda(c)(1 - \mu(b))(C''(c_2, b_2) - C''(c_1, b_2) - C''(c_2, b_1) + C''(c_1, b_1)) \\ + \lambda(c)\mu(d)(C''(c_2, d_2) - C''(c_1, d_2) - C''(c_2, d_1) + C''(c_1, d_1)) \\ + (1 - \lambda(a))(C''(a_2, d_1) - C''(a_1, d_1) - C''(a_2, b_2) + C''(a_1, b_2)) \\ + (1 - \mu(b))(C''(c_1, b_2) - C''(a_2, b_2) - C''(c_1, b_1) + C''(a_2, b_1)) \\ + \lambda(c)(C''(c_2, d_1) - C''(c_1, d_1) - C''(c_2, b_2) + C''(c_1, b_2)) \\ + \mu(d)(C''(c_1, d_2) - C''(a_2, d_2) - C''(c_1, d_1) + C''(a_2, d_1))$$

$$+(C''(c_1, d_1) - C''(a_2, d_1) - C''(c_1, b_2) + C''(a_2, b_2)).$$

donde como vemos, cada sumando consta de 1 a 3 factores, todos ellos positivos por el mismo razonamiento hecho antes. Lo que demuestra que  $C$  es una cópula generada a partir de  $C'$ . ■

Llegamos por fin al teorema tan esperado y que da las claves para entender qué son las cópulas.

**Teorema 1.1** (Teorema de Sklar). *Sea  $H$  una función de distribución conjunta con marginales  $F$  y  $G$  existe una cópula  $C$  cumpliendo*

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

para todo  $x, y \in \overline{\mathbb{R}}$ . Además, si  $F$  y  $G$  son continuas  $C$  es única. Inversamente, si  $F$  y  $G$  son funciones de distribución y  $C$  una cópula, la función  $H$  dada anteriormente es una función de distribución conjunta con marginales  $F$  y  $G$ .

Demostración:

Usando los Lemas 1.1 y 1.2 podemos demostrar la existencia de una subcópula  $C'$  que verifica la igualdad pedida, y extendiendo la subcópula a una cópula queda demostrada la primera parte. En el caso de que las funciones  $F$  y  $G$  sean continuas su rango será el intervalo  $[0, 1]$ , luego la subcópula dada será una cópula, por lo que será única (es al extender la subcópula a la cópula cuando perdemos la unicidad pero como en este caso no hay que hacerlo no se pierde tal unicidad). Finalmente, el recíproco es claro teniendo en cuenta que:

$$H(x, +\infty) = C(F(x), 1) = C(F(x), G(+\infty)) = F(x)$$

$$H(x, -\infty) = C(F(x), 0) = C(F(x), G(-\infty)) = 0.$$

Como vemos en el teorema de Sklar la cópula es la expresión de la distribución conjunta en términos de las marginales. Sin embargo, podemos expresar la cópula en función de la distribución conjunta haciendo uso de la definición de inversar generalizada:

**Definición 1.7.** *Sea  $F$  una función de distribución. La inversa generalizada de  $F$  es cualquier función  $F^{(-1)}$  tal que:*

*Si  $t$  está en  $\text{Ran}F$ , entonces  $F^{(-1)}(t)$  es cualquier número  $x$  en  $\overline{\mathbb{R}}$  tal que  $F(x) = t$ , es decir, para todo  $t$  en  $\text{Ran}F$  se tiene que*

$$F(F^{(-1)}(t)) = t; \tag{1.9}$$

*Si  $t$  está en  $\text{Ran}F$ , entonces:*

$$F^{(-1)}(t) = \inf\{x|F(x) \geq t\} = \sup\{x|F(x) \leq t\}. \tag{1.10}$$

Por tanto, si sustituimos  $x$  por  $F^{(-1)}(u)$  e  $y$  por  $G^{(-1)}(v)$  obtenemos el siguiente corolario:



**Corolario 1.1.1.** Sea  $H$  una función de distribución conjunta,  $F$  y  $G$  sus marginales,  $C'$  la subcópula obtenida mediante el Lema 1.1 y  $F^{(-1)}$  y  $G^{(-1)}$  las inversas generalizadas de  $F$  y  $G$  respectivamente, entonces se tiene que para  $(u, v) \in \text{Dom}C'$ :

$$C'(u, v) = H(F^{(-1)}(u), G^{(-1)}(v)).$$

Además, si  $F$  y  $G$  son continuas  $C'$  es una cópula.

Además, el hecho de que las funciones de distribución caractericen de forma unívoca a las variables aleatorias nos da una segunda interpretación de lo que son las cópulas: la cópula es la expresión en términos de funciones de la dependencia entre las dos variables aleatorias. Por tanto, el teorema de Sklar lo podíamos haber presentado en términos de las variables aleatorias y no de las funciones de distribución asociadas. De ahora en adelante, dado un vector aleatorio  $(X, Y)$  diremos que la cópula que nos da el teorema de Sklar es *la cópula de  $X$  e  $Y$* .

Por otra parte, en un vector aleatorio se dice que sus componentes son independientes si y solo si su función de distribución conjunta es el producto de las distribuciones marginales, es decir,  $H(x, y) = F(x)G(y)$ . Podemos dar un enunciado alternativo de este resultado usando las cópulas como sigue:

**Proposición 1.3.** Sea  $(X, Y)$  un vector aleatorio, sus componentes son independientes si y solo si la cópula de  $X$  e  $Y$  es la cópula  $\Pi$ .

Demostración:

Es claro, ya que usando la definición de la cópula producto  $\Pi$  y el teorema de Sklar se tiene que:

$$H(x, y) = \Pi(F(x), G(y)) = F(x)G(y).$$

■

Este resultado va a ser de especial interés ya que va a establecer, en términos de cópulas, la dependencia de variables aleatorias, además, nos va a dar una idea de “cuán dependientes son las variables aleatorias” dependiendo de “cuanto se aleje la cópula de dichas variables de la cópula producto”.

Finalmente, se puede demostrar usando la regla de la cadena la versión para las densidades del teorema de Sklar.

**Corolario 1.1.2.** Sean  $F, G, H$  y  $C$  verificando el Teorema de Sklar y sean  $f, g, h$  y  $c$  las funciones de densidad y la densidad de la cópula asociadas respectivamente. Entonces se tiene que:

$$h(x, y) = c(F(x), G(y))f(x)g(y).$$

Por lo que la cópula de densidad vendría a estar relacionada con las funciones de densidad de forma similar a como lo hace la cópula con las funciones de distribución.

## 1.3 Medidas de asociación

Como hemos visto, las cópulas nos dan la relación que existe entre dos variables aleatorias pudiendo darnos incluso si dichas variables son independientes o no. Podemos ahora tratar de obtener más información de la dependencia entre dos variables

usando solamente la cópula ya que esta es la que caracteriza dicha dependencia. Hablaremos por tanto de las *medidas de asociación* que serán coeficientes que nos den información sobre la dependencia viniendo determinadas únicamente por la cópula y con independencia de las marginales.

### Concordancia

Informalmente, decimos que un vector aleatorio es concordante cuando grandes valores de una variable están asociados a grandes valores en la otra variable y pequeños valores de una con pequeños valores de la otra. Por ejemplo, los vectores aleatorios con una fuerte correlación lineal directa serán muy concordantes ya que al aumentar una variable la otra también aumentará, mientras que los vectores aleatorios con una fuerte correlación lineal inversa serán muy discordantes ya que a grandes valores de una le corresponderán pequeños valores de la otra. Vamos a ver como se puede expresar esta idea numéricamente.

**Definición 1.8.** Sea  $(X, Y)$  un vector aleatorio de variables aleatorias continuas y sean  $(x_1, y_1)$  y  $(x_2, y_2)$  dos observaciones de dicho vector. Decimos que  $(x_1, y_1)$  y  $(x_2, y_2)$  son concordantes (discordantes) si  $x_1 < x_2$  e  $y_1 < y_2$ , o bien,  $x_1 > x_2$  e  $y_1 > y_2$  ( $x_1 < x_2$  e  $y_1 > y_2$ , o bien,  $x_1 > x_2$  e  $y_1 < y_2$ ). Equivalentemente, dichas medidas son concordantes (discordantes) si  $(x_1 - x_2)(y_1 - y_2) > 0$  ( $(x_1 - x_2)(y_1 - y_2) < 0$ ).

El hecho de que en la definición se usen desigualdades estrictas está justificado porque los vectores los hemos supuesto continuos, luego la probabilidad del punto es 0 por lo que no puedes tener  $x_1 = x_2$  ni  $y_1 = y_2$ . Otra forma alternativa de entender esta definición es la siguiente: dadas dos observaciones y trazada la línea que las une, dichas medidas serán concordantes si la recta tiene pendiente positiva y discordante si es negativa. Vemos que esta definición alternativa no contempla la posibilidad de una pendiente 0 o infinita que son los casos que ya hemos justificado que no se pueden dar.

### Tau de Kendall

En un primer momento vamos a definir la Tau de Kendall en términos de observaciones, como hemos hecho con la concordancia, para llegar más tarde a una definición en función de la cópula, que es nuestro objetivo. Para la definición en función de las observaciones usaremos  $t$  mientras que para la versión poblacional usaremos  $\tau$ .

**Definición 1.9.** Sea  $\{(x_1, x_2), (x_2, y_2), \dots, (x_n, y_n)\}$  una muestra aleatoria de  $n$  observaciones del vector aleatorio continuo  $(X, Y)$ . Por tanto, todas las posibles parejas de dichas observaciones son  $\binom{n}{2}$  de las cuales,  $c$  serán concordantes y  $d$  serán discordantes según la definición anterior. Definimos entonces la Tau de Kendall para dicha muestra como

$$t = \frac{c - d}{c + d} = \frac{c - d}{\binom{n}{2}}.$$

Es decir,  $t$  es la probabilidad de la concordancia menos la discordancia entre un par de observaciones. Podemos entonces formular la Tau de Kendall en su versión poblacional como sigue.

**Definición 1.10.** Sean  $(X_1, Y_1)$  y  $(X_2, Y_2)$  dos vectores aleatorios continuos independientes e idénticamente distribuidos. Definimos entonces la Tau de Kendall para el vector aleatorio  $(X, Y)$  como:

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

El siguiente teorema relaciona la Tau de Kendall con las cópulas dando una tercera definición a este coeficiente.

**Teorema 1.2.** Sean  $(X_1, Y_1)$  y  $(X_2, Y_2)$  dos vectores aleatorios independientes e idénticamente distribuidos cuyas marginales son  $F$  y  $G$  y cuya cópula asociada es  $C$ . Entonces se tiene que:

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

Demostración:

Como los vectores aleatorios son continuos tenemos que  $\mathbb{P}[(X_1 - X_2)(Y_1 - Y_2) < 0] = 1 - \mathbb{P}[(X_1 - X_2)(Y_1 - Y_2) > 0]$ , por lo que sustituyendo en nuestra definición poblacional de la Tau de Kendall obtenemos:

$$\tau = 2\mathbb{P}[(X_1 - X_2)(Y_1 - Y_2) > 0] - 1.$$

Partiendo esta probabilidad en dos (tal y como definimos al principio la concordancia) podemos usar que  $\mathbb{P}[(X_1 - X_2)(Y_1 - Y_2) > 0] = \mathbb{P}[X_1 < X_2, Y_1 < Y_2] + \mathbb{P}[X_1 > X_2, Y_1 > Y_2]$ . Si nos fijamos en cualquiera de los sumandos vemos que:

$$\mathbb{P}[X_1 < X_2, Y_1 < Y_2] = \int \int_{\mathbb{R}^2} C(F(x), G(y)) dC(F(x), G(y)).$$

Y transformando las probabilidades como  $u = F(x)$  y  $v = G(y)$  llegamos a:

$$\mathbb{P}[X_1 < X_2, Y_1 < Y_2] = \int \int_{\mathbb{I}^2} C(u, v) dC(u, v).$$

Finalmente:

$$\begin{aligned} \tau &= 2\mathbb{P}[(X_1 - X_2)(Y_1 - Y_2) > 0] - 1 = 2(\mathbb{P}[X_1 < X_2, Y_1 < Y_2] + \mathbb{P}[X_1 > X_2, Y_1 > Y_2]) - 1 \\ &= 2(2 \int \int_{\mathbb{I}^2} C(u, v) dC(u, v)) - 1 = 4 \int \int_{\mathbb{I}^2} C(u, v) dC(u, v) - 1. \end{aligned}$$

Lo que concluye la demostración. ■

Aunque, la fórmula del teorema 1.2 no siempre es fácil de computar, ya que  $C$  puede ser singular o tener una componente singular. El resultado que probamos a continuación brinda una forma más fácil de calcular la Tau de Kendall para estos casos.

**Teorema 1.3.** Sea  $C$  una cópula tal que el producto  $(\partial C / \partial u)(\partial C / \partial v)$  es integrable en  $\mathbb{I}^2$ , entonces:

$$\int \int_{\mathbb{I}^2} C(u, v) dC(u, v) = \frac{1}{2} - \int \int_{\mathbb{I}^2} \frac{\partial}{\partial u} C(u, v) \frac{\partial}{\partial v} C(u, v) du dv.$$

Demostración:

En el mejor de los casos  $C$  es absolutamente continua, por lo que podemos proceder paso a paso.

$$\int \int_{\mathbb{I}^2} C(u, v) dC(u, v) = \int_0^1 \int_0^1 \frac{\partial^2 C(u, v)}{\partial u \partial v} du dv.$$

Integrando por partes respecto de  $u$  obtenemos:

$$\begin{aligned} \int_0^1 C(u, v) \frac{\partial^2 C(u, v)}{\partial u \partial v} du &= [C(u, v) \frac{\partial C(u, v)}{\partial u}]_{u=0}^{u=1} - \int_0^1 \frac{\partial C(u, v)}{\partial u} \frac{\partial C(u, v)}{\partial v} du \\ &= v - \int_0^1 \frac{\partial C(u, v)}{\partial u} \frac{\partial C(u, v)}{\partial v} du. \end{aligned}$$

Podemos separar este resultado para hacer la segunda integral obteniendo  $\int_0^1 v dv = \frac{1}{2}$  por lo que se verifica el resultado ya que la integral en  $v$  de la integral en  $u$  da la doble integral buscada.

Supongamos ahora el caso en el que  $C$  tiene una componente singular. Sea entonces  $0 = u_0 < u_1 < \dots < u_n = 1$  una partición en el eje  $u$  y  $0 = v_0 < v_1 < \dots < v_m = 1$  una partición en el eje  $v$  con  $\Delta u_i = u_i - u_{i-1}$  y  $\Delta v_i = v_i - v_{i-1}$ . Podemos entonces hacer una partición de  $\mathbb{I}^2$  en rectángulos de la forma  $R_{ij} = [u_{i-1}, u_i] \times [v_{j-1}, v_j]$ . Utilizando la medida de estos rectángulos definida podemos describir la integral como una suma de Riemann:

$$\sum_{j=1}^m \sum_{i=1}^n C(u_i, v_j) [C(u_i, v_j) - C(u_{i-1}, v_j) - C(u_i, v_{j-1}) + C(u_{i-1}, v_{j-1})].$$

Si partimos esta suma convenientemente conseguimos estas otras dos:

$$\begin{aligned} \sum_{j=1}^m \sum_{i=1}^n [(C(u_i, v_j))^2 - C(u_i, v_j)C(u_i, v_{j-1}) - (C(u_{i-1}, v_j))^2 + C(u_{i-1}, v_j)C(u_{i-1}, v_{j-1})] \\ + \sum_{j=1}^m \sum_{i=1}^n [C(u_i, v_j) - C(u_{i-1}, v_j)][C(u_{i-1}, v_j)C(u_{i-1}, v_{j-1})]. \end{aligned}$$

Donde el primer sumando es una serie telescópica respecto a la  $i$ , por lo que podemos calcularla y dejarla como una sola suma con índice  $j$  en lugar de una doble suma.

$$\begin{aligned} \sum_{j=1}^m [(C(1, v_j))^2 - C(1, v_j)C(1, v_{j-1}) - (C(0, v_j))^2 + C(0, v_j)C(0, v_{j-1})] \\ = \sum_{j=1}^m [v_j^2 - v_j v_{j-1}] = \sum_{j=1}^m v_j \Delta v_j. \end{aligned}$$

Que es una suma de Riemann para la integral  $\int_0^1 v dv = \frac{1}{2}$ . Por otra parte, la otra suma la podemos reescribir para que también sea una suma de Riemann como:

$$\sum_{j=1}^m \sum_{i=1}^n \left[ \frac{C(u_i, v_j) - C(u_{i-1}, v_j)}{\Delta u_i} \right] \left[ \frac{C(u_{i-1}, v_j) C(u_{i-1}, v_{j-1})}{\Delta v_j} \right] \Delta u_i \Delta v_j.$$

Obteniendo la integral que nos faltaba para completar la igualdad de la fórmula. ■

Si aplicamos ahora los dos teoremas anteriores obtenemos el siguiente resultado, que será el que nos permita calcular la Tau de Kendall más adelante con facilidad.

**Corolario 1.3.1.** *Sea  $C$  una cópula tal que el producto  $(\partial C / \partial u)(\partial C / \partial v)$  es integrable en  $\mathbb{I}^2$ , entonces:*

$$\tau = 1 - 4 \int \int_{\mathbb{I}^2} \frac{\partial}{\partial u} C(u, v) \frac{\partial}{\partial v} C(u, v) du dv.$$

### Rho de Spearman

Otra medida de asociación muy usada es la Rho de Spearman. En este caso, también se define en términos de la concordancia y discordancia de vectores aleatorios, pero esta vez los vectores son  $(X_1, Y_1)$  y  $(X_2, Y_3)$ , de forma que la pareja  $X_1$  y  $X_2$  y la pareja  $Y_1$  e  $Y_3$  tienen las mismas distribuciones marginales  $F$  y  $G$  respectivamente pero las cópulas son:  $C$  para el primer vector y  $\Pi$  para el segundo vector. Esto viene de tomar 3 vectores aleatorios continuos independientes e idénticamente distribuidos  $(X_1, Y_1)$ ,  $(X_2, Y_2)$  y  $(X_3, Y_3)$  por lo que todas las  $X_n$  comparten la marginal  $F$ , todas las  $Y_n$  comparten la marginal  $G$ , la cópula de cualquier vector  $(X_i, Y_i)$  es  $C$  y la cópula de cualquier vector  $(X_i, Y_j)$  con  $i \neq j$  es la cópula producto  $\Pi$  por ser independientes entre ellos.

**Definición 1.11.** *Sean  $(X_1, Y_1)$ ,  $(X_2, Y_2)$  y  $(X_3, Y_3)$  tres vectores aleatorios continuos, independientes e idénticamente distribuidos. Definimos entonces la Rho de Spearman para el vector aleatorio  $(X, Y)$  como:*

$$\rho = 3(\mathbb{P}(X_1 - X_2)(Y_1 - Y_3) > 0] - \mathbb{P}(X_1 - X_2)(Y_1 - Y_3) < 0]).$$

Se puede usar también en la definición el vector  $(X_3, Y_2)$  en lugar de  $(X_2, Y_3)$  obteniendo lo mismo ya que los vectores están idénticamente distribuidos. El 3 que aparece como coeficiente es una constante de normalización para que  $\rho \in [-1, 1]$ .

Introducimos ahora el teorema que relaciona esta medida de asociación con la cópula para luego hablar de esta relación.

**Teorema 1.4.** *Sean  $(X_1, Y_1)$ ,  $(X_2, Y_2)$  y  $(X_3, Y_3)$  tres vectores aleatorios continuos, independientes e idénticamente distribuidos cuyas marginales son  $F$  y  $G$  y cuya cópula asociada es  $C$ . Entonces se tiene que:*

$$\rho = 12 \int \int_{\mathbb{I}^2} uv dC(u, v) - 3 = 12 \int \int_{\mathbb{I}^2} C(u, v) du dv - 3.$$

Demostración:

Comenzamos usando la probabilidad del contrario para obtener 2 veces la misma probabilidad de la concordancia menos 1:

$$3(\mathbb{P}[(X_1 - X_2)(Y_1 - Y_3) > 0] - \mathbb{P}[(X_1 - X_2)(Y_1 - Y_3) < 0]) = 6\mathbb{P}[(X_1 - X_2)(Y_1 - Y_3) > 0] - 3.$$

Y si aplicamos como antes que la probabilidad del producto la podemos expresar como dos probabilidades nos da que:

$$6[\mathbb{P}[X_1 < X_2, Y_1 < Y_3] + \mathbb{P}[X_2 < X_1, Y_3 < Y_1]] - 3.$$

Estas probabilidades las podemos expresar como integrales respecto a  $\Pi(F(x), G(y))$  o respecto a  $C(F(x), G(y))$ . El procedimiento en ambos casos es el mismo obteniendo en uno la primera igualdad y en el otro la segunda. Vamos a hallar la segunda aunque como hemos dicho, la primera se halla de la misma forma:

$$\begin{aligned} & 6[\mathbb{P}[X_1 < X_2, Y_1 < Y_3] + \mathbb{P}[X_2 < X_1, Y_3 < Y_1]] - 3 \\ &= 6 \left[ \int \int_{\mathbb{R}^2} \mathbb{P}[X_1 < x, Y_1 < y] d\Pi(F(x), G(y)) + \int \int_{\mathbb{R}^2} \mathbb{P}[X_1 > x, Y_1 > y] d\Pi(F(x), G(y)) \right] - 3 \\ &= 6 \left[ \int \int_{\mathbb{R}^2} C(F(x), G(y)) d\Pi(F(x), G(y)) + \int \int_{\mathbb{R}^2} (1 - F(x) - G(y) + C(F(x), G(y))) d\Pi(F(x), G(y)) \right] - 3 \\ &= 6 \left[ \int \int_{\mathbb{I}^2} C(u, v) dudv + \int \int_{\mathbb{I}^2} (1 - u - v + C(u, v)) dudv \right] - 3 \\ &= 6 \left[ \int \int_{\mathbb{I}^2} C(u, v) dudv + 1 - \frac{1}{2} - \frac{1}{2} \int \int_{\mathbb{I}^2} C(u, v) dudv \right] - 3 \\ &= 12 \int \int_{\mathbb{I}^2} C(u, v) dudv - 3. \end{aligned}$$

Lo que concluye la demostración. ■

Sin embargo, aún no sabemos qué significa la Rho de Spearman de una cópula realmente. Para poder llegar a una interpretación de esta medida de asociación vamos a partir de la segunda expresión que hemos dado y vamos a operar como sigue.

$$\begin{aligned} \rho &= 12 \int \int_{\mathbb{I}^2} C(u, v) dudv - 3 \\ &= 12 \int \int_{\mathbb{I}^2} C(u, v) dudv - \int \int_{\mathbb{I}^2} uv dudv \\ &= 12 \int \int_{\mathbb{I}^2} [C(u, v) - uv] dudv \\ &= 12 \int \int_{\mathbb{I}^2} [C(u, v) - \Pi(u, v)] dudv. \end{aligned}$$

Luego vemos que, la Rho de Spearman es la "distancia media", multiplicada por 12 como coeficiente, entre la cópula dada y la cópula producto. Dicho de otro modo, la Rho de Spearman da una medida de cuan lejos está nuestro vector aleatorio de tener variables independientes.

### Dependencia de colas

Las medidas de dependencia anteriores han ido siempre ligadas al concepto de concordancia y discordancia. Esto quiere decir que las medidas anteriores tenían en cuenta como de grandes o pequeñas eran unas variables cuando las otras lo eran. La dependencia de colas también mide esto último pero sin recurrir al concepto de concordancia ni de discordancia.

**Definición 1.12.** Sean  $X$  e  $Y$  variables aleatorias continuas con distribuciones  $F$  y  $G$  respectivamente. La dependencia de la cola superior  $\lambda_U$  es el límite, si existe, de la probabilidad condicional de que  $Y$  supere la curva de nivel  $t$  dado que  $X$  la supera cuando la curva de nivel tiende a 1, es decir:

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

De igual forma, la dependencia de la cola inferior  $\lambda_L$  es el límite, si existe, de la probabilidad condicional de que  $Y$  esté por debajo de la curva de nivel  $t$  dado que  $X$  está por debajo cuando la curva de nivel tiende a 0, es decir:

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

Pero esta definición no es del todo manejable y además no se aprecia que dependa exclusivamente de la cópula. Por tanto, echamos mano del siguiente teorema.

**Teorema 1.5.** Sean  $X$ ,  $Y$ ,  $F$ ,  $G$ ,  $\lambda_U$  y  $\lambda_L$  como en la definición anterior y  $C$  la cópula de  $X$  e  $Y$ . Si  $\lambda_U$  y  $\lambda_L$  existen, entonces:

$$\lambda_U = \lim_{t \rightarrow 1^-} \frac{1 - 2t + C(t, t)}{1 - t} \quad (1.11)$$

$$\lambda_L = \lim_{t \rightarrow 0^+} \frac{C(t, t)}{t}. \quad (1.12)$$

#### Demostración:

Demostramos solo la primera igualdad, la otra se hace de forma similar.

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

■

## 1.4 Familias de cópulas

Las familias de cópulas son conjuntos de estas, en los cuales, las cópulas que los componen se asemejan en cuanto a su forma. La más interesante es la cópula empírica al permitir dar una cópula, y por tanto una relación de dependencia, para cualquier muestra. Aunque esta dependencia no se acerque demasiado a la real es muy útil a la hora de modelizar dicha dependencia por su parecido con otras cópulas. También damos la fórmula de otras familias de cópulas donde los miembros solo se diferencian en uno o varios parámetros. El objetivo de estas cópulas normalmente será parecerse lo máximo posible a la cópula empírica por medio de ese parámetro.

### La cópula empírica

El objetivo de esta subsección es crear una cópula a partir de un conjunto de observaciones. Pero con esta cópula no podremos usar las fórmulas estudiadas para la Tau de Kendall ya que esta cópula no va a ser derivable. Por lo que proporcionaremos un teorema para calcularlas aunque la demostración se puede encontrar en [7] y en [2].

**Definición 1.13.** Sean  $\{(x_k, y_k)\}_{k=1}^n$  una muestra de tamaño  $n$  de una distribución bivalente continua. Llamamos cópula empírica a la cópula dada por:

$$C_n\left(\frac{i}{n}, \frac{j}{n}\right) = \frac{\text{número de parejas } (x, y) \text{ en la muestra tales que } x \leq x_i \text{ e } y \leq y_j}{n} \quad (1.13)$$

donde  $x_i$  e  $y_j$  con  $1 \leq i, j \leq n$  son los ordinales de la muestra.

La densidad de la cópula empírica estaría por tanto dada por:

$$c_n\left(\frac{i}{n}, \frac{j}{n}\right) = \begin{cases} 1/n & \text{si } (x_i, y_i) \text{ es un elemento de la muestra} \\ 0 & \text{si no} \end{cases} \quad (1.14)$$

Y ambas cópulas se pueden relacionar por:

$$C_n\left(\frac{i}{n}, \frac{j}{n}\right) = \sum_{p=1}^i \sum_{q=1}^j c_n\left(\frac{p}{n}, \frac{q}{n}\right).$$

Lo que tiene sentido, ya que la cópula funcionaría como la función de frecuencias acumuladas de la densidad de la cópula así como la función de distribución es la integral o la suma acumulativa de la función de densidad.

Finalmente, para expresar las medidas de asociación de la cópula empírica tenemos el siguiente teorema, cuya demostración puede encontrarse en [7] y en [2].

**Teorema 1.6.** Sea  $C_n$  la cópula empírica definida según 1.13 y  $c_n$  la cópula de densidad empírica de 1.14. Sus medidas de asociación verifican:

$$\tau = \frac{2n}{n-1} \sum_{i=2}^n \sum_{j=2}^n \sum_{p=1}^{i-1} \sum_{q=1}^{j-1} \left[ c_n\left(\frac{i}{n}, \frac{j}{n}\right) c_n\left(\frac{p}{n}, \frac{q}{n}\right) - c_n\left(\frac{i}{n}, \frac{q}{n}\right) c_n\left(\frac{p}{n}, \frac{j}{n}\right) \right] \quad (1.15)$$



$$\rho = \frac{12}{n^2 - 1} \sum_{i=1}^n \sum_{j=1}^n \left[ C_n \left( \frac{i}{n}, \frac{j}{n} \right) - \frac{i}{n} \cdot \frac{j}{n} \right] \quad (1.16)$$

$$\lambda_U = 2 - 2 \exp \left\{ \frac{1}{n} \sum_{i=1}^n \ln \left( \sqrt{\ln \left( \frac{n+1}{R_i} \right) \ln \left( \frac{n+1}{S_i} \right)} / \ln \left( 1 / \max \left( \frac{R_i}{n+1}, \frac{S_i}{n+1} \right)^2 \right) \right) \right\}. \quad (1.17)$$

### Cópulas paramétricas

A pesar de haber muchas más familias de cópulas paramétricas las más importantes y las más usadas son las que están listadas en la Tabla 1.1.

Nombre	Cópula	Rango
Ali-Mikhail-Haq(AMH)	$C(u, v) = (1 - \alpha) \left[ \left( \frac{1 - \alpha}{u} \right) \left( \frac{1 - \alpha}{v} \right) - \alpha \right]^{-1}$	$\alpha \in [-1, 1]$
Clayton	$C(u, v) = [u^{-\alpha} + v^{-\alpha} - 1]^{-1/\alpha}$	$\alpha \in [0, +\infty)$
Farlie-Gumbel-Morgenstein(FGM)	$C(u, v) = uv[1 + \phi(1 - u)(1 - v)]$	$\phi \in [-1, 1]$
Frank	$C(u, v) = \log_{\alpha} \left[ 1 + \frac{(\alpha^u - 1)(\alpha^v - 1)}{\alpha - 1} \right]$	$\alpha \in [0, +\infty)$
Galambos	$C(u, v) = uv \exp \left\{ [ (1 - u)^{-\theta} + (1 - v)^{-\theta} ]^{-1/\theta} \right\}$	$\theta \in [0, +\infty)$
Gumbel	$C(u, v) = \exp \left\{ - \left[ (-\log u)^{\phi} + (-\log v)^{\phi} \right]^{1/\phi} \right\}$	$\phi \in [1, +\infty)$
Hüsler-Reiss	$C(u, v) = \exp \left[ \Phi \left\{ \frac{a}{2} + \frac{1}{a} \log \frac{\log u}{\log v} \right\} \log u + \Phi \left\{ \frac{a}{2} + \frac{1}{a} \log \frac{\log v}{\log u} \right\} \log v \right]$	$a \in [0, +\infty)$
Normal	$C(u, v) = \Phi_{\Sigma}(\Phi^{-1}u, \Phi^{-1}v)$	$\Sigma$ matriz de covarianzas
Plackett	$C(u, v) = \frac{1 + (\theta - 1)(u + v) - \sqrt{[1 - (\theta - 1)(u + v)]^2 - 4\theta(\theta - 1)uv}}{2(\theta - 1)}$	$\theta \in [0, +\infty)$
Student	$C(u, v) = \int_0^1 \Phi_{\Sigma}(x, y) ds$	$\Sigma$ matriz de covarianzas
Tawn	$C(u, v) = \exp \left\{ \log uv \left( 1 - \theta \frac{\log v}{\log uv} + \theta \left( \frac{\log v}{\log uv} \right)^2 \right) \right\}$	$\theta \in [0, 1]$

Tabla 1.1: Familias de cópulas paramétricas

Con  $\Phi$  la función de distribución de una normal estándar y  $\Phi_{\Sigma}$  la función de distribución de una normal bivalente con vector de medias nulo y matriz de covarianzas  $\Sigma$ . En el caso de la Student,  $x = t_{v_1}^{-1}(u)/G_{v_1}^{-1}(s)$  e  $y = t_{v_1}^{-1}(v)/G_{v_1}^{-1}(s)$  siendo  $t_v$  la función de distribución de una Student con  $v$  grados de libertad y siendo  $G_v$  la función de distribución de la distribución  $\sqrt{v/\chi_v^2}$ . Es usual que cuando hay una matriz de covarianzas se asuma que las varianzas son 0 dejando como único parámetro la covarianza, esto se aplica a la cópula normal y de Student, además en el caso de esta última también se asume que todos los grados de libertad son el mismo.



## Aplicación en finanzas

### Introducción

Como sabemos, las cópulas son una herramienta bastante útil a la hora de modelizar dependencia. En este capítulo estudiamos como se han usado en el artículo de Grégoire [4]. Por otra parte, en la industria de la energía se ve que el precio de un tipo de energía afecta a otros tipos, en nuestro caso el del gas natural y del petróleo. Esta dependencia se ha probado que no se puede modelizar con una normal bivalente ni con regresiones lineales tras aplicar el logaritmo. Al ser la cópula la expresión en sí de la dependencia, trataremos de encontrar la que mejor se ajuste para así poder hablar de dicha dependencia.

Para facilitar los cálculos trabajaremos con el logaritmo del rendimiento, entendiendo el rendimiento como  $r_i = \frac{p_i}{p_j} - 1$  con  $i < j$ , la justificación de esto la podemos encontrar en [8]. En el texto se plantean varias familias de cópulas para describir la dependencia mencionada. De entre ellas se escoge el mejor representante minimizando el error. Para esta elección se usa un test de ajuste de bondad y una vez elegida la cópula se realizarán predicciones usando simulaciones de Monte Carlo. La posible existencia de autocorrelación y heterocedasticidad es también tenida en cuenta y trabajada a lo largo del texto.

Los datos que se usaron para este estudio fueron obtenidos del servicio de información financiera Bloomberg y son los precios del gas natural y del petróleo basados en los contratos en Nymex(New York Mercantile Exchange).

### Estimación de las marginales

Lo primero que se realiza son los test Box-Pierce y Ljung-Box para los logaritmos de ambas series temporales para detectar la existencia de autocorrelación y heterocedasticidad.

En el caso del petróleo ninguno de estos test es significativo al nivel del 5%, por lo que, las muestras del petróleo  $O_1, O_2, \dots$  pueden entenderse como muestras de una distribución  $F$  con media  $\mu = 1,220 \times 10^{-3}$  y desviación estándar  $\sigma = 1,912 \times 10^{-3}$ . Sin embargo, como las colas de esta distribución demuestran ser mayores que las de la normal, el error  $X_j = (O_j - \mu)/\sigma$  no es normal tal y como indica el test de ajuste de bondad. Por otra parte, con el mismo test se puede ver que la distribución de Student con  $\Psi = 13,745$  grados de libertad se ajusta mucho mejor.

No obstante, al realizarle los tests anteriores al gas natural se nota que sí existe una autocorrelación con varios desfases. Hay entonces que desechar la hipótesis nula de homocedasticidad al nivel del 5% para el desfase de 1 mes y al nivel del 0,1% para el desfase de 10 meses. Por tanto, es necesario utilizar el método Garch(heterocedasticidad condicional autoregresiva generalizada). Se obtiene así que la serie referente al rendimiento del gas natural se asemeja al siguiente modelo:

$$N_t = 1,281 \times 10^{-4} + Y_t h_t^{1/2}$$

$$h_t = 2,138 \times 10^{-4} + 6,206 \times 10^{-2} N_{t-1}^2$$

$$+ 3,663 \times 10^{-3} N_{t-2}^2 + 5,862 \times 10^{-2} N_{t-4}^2$$

$$+9,319 \times 10^{-2} N_{t-5}^2 + 9,472 \times 10^{-2} N_{t-6}^2 + 0,535 h_{t-1}.$$

En este caso los errores siguen una distribución de Student asimétrica  $Y_t = \xi + V^{-1/2}Z$  con  $\xi = -0,622$  como parámetro de localización,  $V$  como una distribución chi cuadrado con 6,737 grados de libertad y  $Z$  como una normal asimétrica independiente de  $V$  con parámetros  $\Omega = 0,986$  y  $\alpha = 1,014$ .

### *Estudio de la dependencia*

Una vez están descritos los movimientos diarios en los precios del petróleo y del gas natural es necesario estudiar la dependencia entre estos. Existen dos opciones: que sean independientes, luego la cópula asociada sería la producto en virtud de la proposición 1.3, o que sean dependientes, luego habría que encontrar otra cópula para modelizar esta dependencia existente y conseguir una función de distribución conjunta.

En la Figura 2.1 encontramos todos los puntos de este estudio normalizados por rangos. En ella, se observa que parece haber una dependencia positiva entre ambas variables, es decir, hay indicio de que la cópula buscada no va a ser la cópula producto.

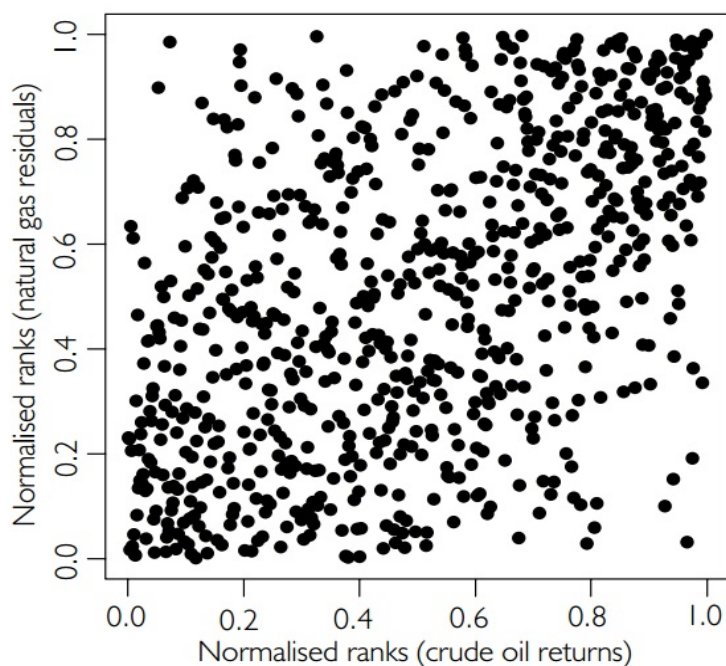


Figura 2.1: Observaciones de los precios del petróleo y el gas natural normalizadas por rangos

Para desechar la hipótesis nula de que la cópula es la producto se calcula la cópula empírica  $C_n$  según 1.13 y se obtienen tanto su Rho de Spearman como su Tau de Kendall con 1.15 y 1.16. Esto da unos valores de  $\rho_n = 0,508$  y  $\tau_n = 0,349$  que se alejan bastante del cero con un valor  $P < 0,0001$  por lo que la hipótesis de que la cópula es la producto es rechazada.

## Selección de la cópula: estimación

Una vez queda probado que no existe independencia entre las variables, es hora de encontrar una cópula  $C$  que describa esta dependencia. Para ello, se estima la cópula por sus medidas de dependencia, la Tau de Kendall y la Rho de Spearman. Sin embargo, en esta investigación, solo es usada la Tau de Kendall para la estimación ya que es más fácil de calcular, a pesar de que ambos estimadores son válidos.

El método descrito a continuación es el denominado “Inversión de la Tau de Kendall”. Este consiste en calcular la Tau de Kendall para unas determinadas familias de cópulas paramétricas obteniendo una función  $\tau(\theta)$ , y, al igualarla a la  $\tau_n$  calculada empíricamente obtener el valor de  $\theta_n$  que tendría la cópula que mejor se adapte de cada determinada familia. Es decir, consiste en elegir de cada familia la cópula cuya Tau de Kendall coincida con la de la cópula empírica en caso de ser posible. No obstante, podría no ser posible siempre debido a restricciones en el rango de los parámetros en la familia de cópulas, en cuyo caso, se desecharían las cópulas de esa familia como posibles cópulas solución. En la Tabla 2.1 se especifica la función inversa a la  $\tau(\theta)$  y el valor de  $\theta_n$  en función de la  $\tau_n = 0,349$ .

Familia	Fórmula	$\theta_n$
Clayton	$\theta = 2\tau/(1 - \tau)$ para $\tau > 0$	1,076
Frank	Obtenida numéricamente	3,474
Gumbel	$\theta = 1/(1 - \tau)$ para $\tau > 0$	1,538
Plackett	Obtenida numéricamente	5,107
Normal	$\theta = \sin(\pi\tau/2)$	0,522
Student	$\theta = \sin(\pi\tau/2)$	0,522

Tabla 2.1: Estimaciones del parámetro  $\theta$  para las cópulas

Resaltamos que la cópula de Student es biparamétrica, pero en el cálculo de la  $\tau$  el parámetro  $\gamma$  no interviene así que simplemente se la  $\theta$  y se deja la  $\gamma$  libre para ajustarla más adelante según sea requerido.

## Selección de la cópula: test de ajuste

Por lo pronto las cópulas candidatas a solución solo se asemejan a la cópula empírica en la Tau de Kendall, pero esto no es suficiente. Para ello se lleva a cabo un test de bondad de ajuste usando el método bootstrap y el estadístico Cramér-von Mises definido como:

$$S_n = n \sum_{t=1}^n \{C_{\theta_n}(u_t, v_t) - C_n(u_t, v_t)\}^2.$$

Donde recordemos  $C_n$  es la cópula empírica y  $C_{\theta_n}$  es la cópula paramétrica con el  $\theta$  calculado según la inversión de la Tau de Kendall. El método bootstrap consiste en: generar una pareja de valores aleatorios  $(u_i, t_i)$  de una distribución uniforme  $U(0, 1)$  y dados estos valores, hallar la  $v_i$  que acompañaría a dicha  $u_i$  como  $C(u_i, v_i) = u_i t_i$  despejando  $v_i$  de dicha ecuación. Se obtiene así un conjunto de pseudo-observaciones  $\{(u_i, v_i)\}_{i=1}^N$ , del cual, es posible hallar una cópula empírica  $C_m^*$ . Para dicha cópula empírica se puede obtener su correspondiente  $\theta_m^*$  repitiendo lo anterior y, con la fórmula

anterior,  $S_m^*$ . Para más información referente a este método así como su validez se puede consultar [3]. En el estudio, se generaron 2.000 muestras bootstrap  $\{S_1^*, S_2^*, \dots, S_{2000}^*\}$  de esta forma y los valores  $P$  se obtuvieron según:

$$P = \frac{1}{2000} \sum_{m=1}^{2000} H(S_m^* - S_n).$$

Donde  $H$  es la función escalón de Heavyside o función escalón unitario. Haciendo este proceso para las seis cópulas consideradas se consiguieron los valores  $P$  de la Tabla 2.2.

Familia	Valor $P$
Clayton	0,0000
Frank	0,0095
Gumbel	0,0075
Plackett	0,0080
Normal	0,0140
Student, $\gamma = 22$	0,0300

Tabla 2.2: Valores  $P$  para los ajustes de las cópulas

Como vemos, la cópula que mayor valor  $P$  da es la Student y solo si los grados de libertad  $\gamma$  son 22. Por otra parte, la cópula normal también se ajusta de forma aceptable con un valor  $P$  de 1,4%, pero existen varios problemas con la cópula normal: la cópula normal exige una normalidad en las marginales de forma que la distribución conjunta sería una Normal bivalente, cosa que se ha especificado que no se satisface; la cópula Normal no pone de manifiesto los comportamientos en las colas; y la cópula Normal forzaría una dependencia lineal en las variables, cosas que no existe.

Finalmente, podemos contruir la función de distribución conjunta como:

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

Donde  $F$  es la función de distribución de una Student,  $G$  es la función de distribución de la Student asimétrica vista antes y  $C$  es la cópula de Student con parámetro  $\theta = 0,522$  y  $\gamma = 22$  grados de libertad.

### *Haciendo predicciones*

Para hacer predicciones es necesario generar unos valores  $(u_t, v_t)$  dado un tiempo  $t$ . La forma de generarlos es idéntica a como se hace en el método bootstrap, pero esta vez  $t$  es el tiempo, no el valor de la cópula. Una vez tenemos estos valores se tienen que hallar las preimágenes de dichos valores según  $X_t = F^{(-1)}(u_t)$  y  $Y_t = G^{(-1)}(v_t)$  siendo  $F^{(-1)}$  y  $G^{(-1)}$  las inversas generalizadas de  $F$  y  $G$ , las funciones de distribución con las que hemos decidido modelizar los precios del petróleo y el gas natural. Finalmente, se calculan los precios usando  $O_t = \sigma X_t + \mu$  para el del petróleo y el método Garch antes mencionado a la inversa para hallar el del gas natural  $N_t$ . Repitiendo este proceso para distintos  $t$  es posible obtener series temporales para ambos precios. No obstante, siempre será necesario repetir el proceso muchas veces para obtener una mayor fiabilidad del resultado obtenido.



## Resultados simulados

Usando todo lo anterior, se hizo una simulación de 4,000,000 caminos aleatorios de un año de duración en las series temporales empezando el 19 de julio de 2006 obteniendo la Figura 2.2 con la distribución de los precios para julio de 2007. En ella, vemos una clara asimetría y un gran peso en las colas, lo que justifica muchas de las elecciones tomadas anteriormente. Es posible ahora generar distribuciones de probabilidad condicionadas fijada una de las variables para hacer predicciones sobre la otra. Por ejemplo, en la Figura 2.3 vemos que supuesto que el precio del gas ronda los 70\$ lo más probable es que el precio del petróleo ronde los 6,75\$.

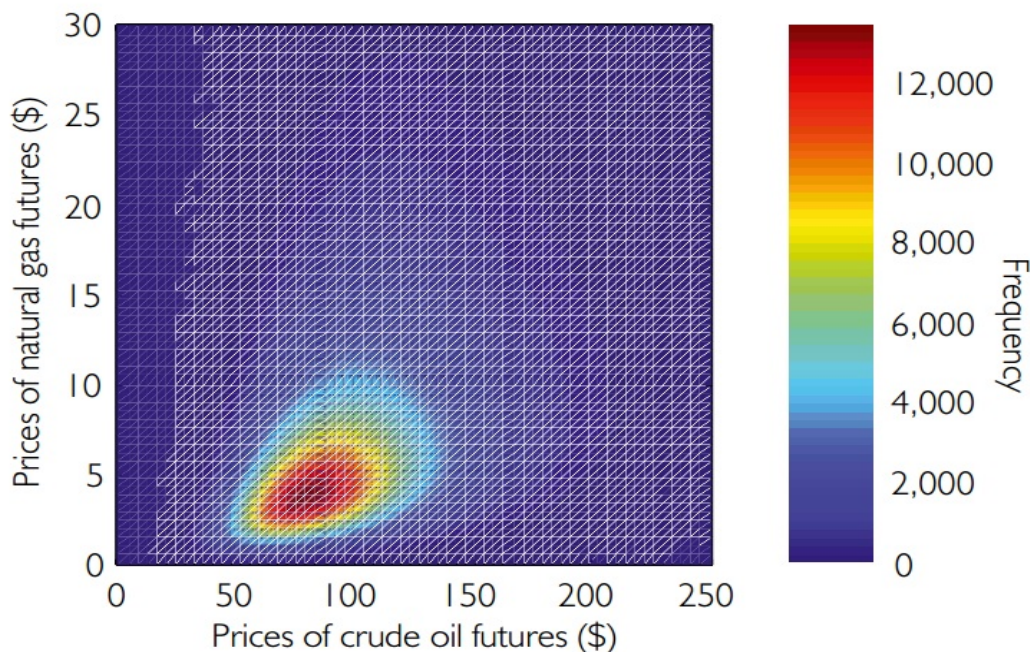


Figura 2.2: Predicción de la distribución de los posibles precios del petróleo y del gas natural en julio de 2007

## Otras consideraciones

A lo largo de toda la discusión se ha supuesto que la dependencia entre el precio del gas natural y del petróleo es constante respecto al tiempo. Es decir, que la cópula que relaciona ambos precios no varía en función del tiempo. Sin embargo, al calcular los  $\tau(t)$  referentes a 250 días se observa que esta  $\tau$  no es constante, aunque, a partir de abril de 2004 este valor empieza a estabilizarse en torno al 0,35. Entendemos entonces que todo lo anterior es correcto debido a la estabilidad que toma la  $\tau$  a partir de 2004 ya que la podemos asumir constante de ahí en adelante.

Por otra parte, podría darse el peor de los casos, que esta dependencia no se mantuviese constante ni se estabilizase en ningún momento. En cuyo caso, tendríamos que buscar una cópula  $C_{\theta_t}$  para modelizar la dependencia que variaría en función de  $\theta_t$ , la cual tendríamos que asumir que varía en función del tiempo de alguna forma conocida, como por ejemplo linealmente  $\theta_t = at + b$  para simplificar cálculos. En este caso,

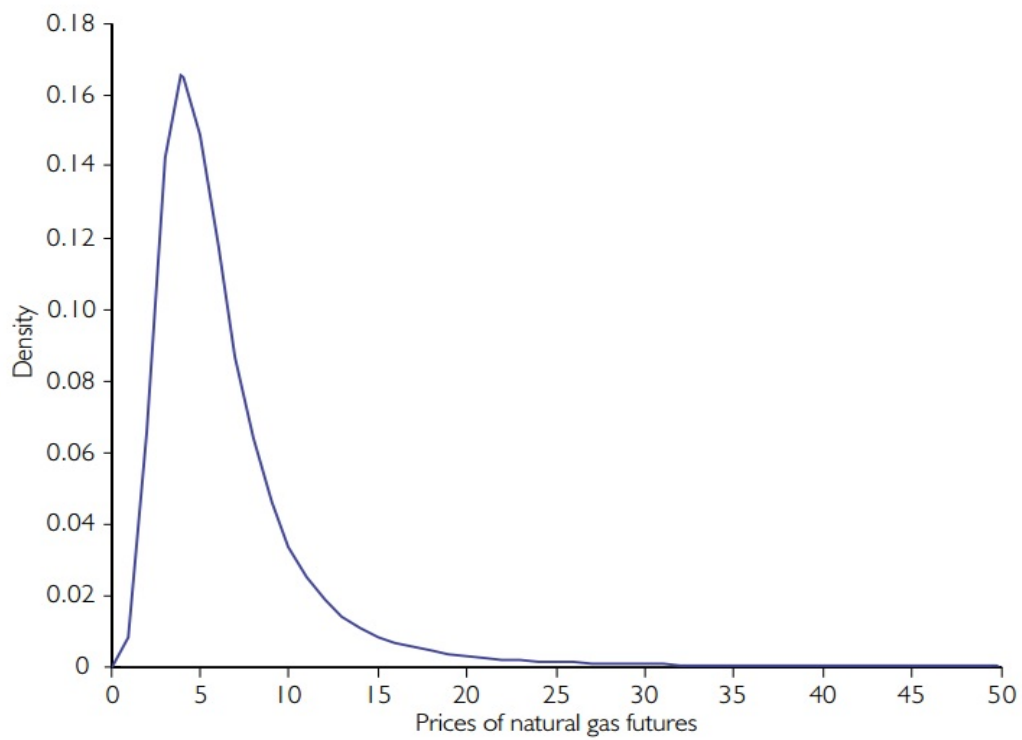


Figura 2.3: Predicción de la distribución del posible precio del gas natural en julio de 2007 si el petróleo cuesta 70 \$

hablaríamos de cópulas dinámicas sobre las cuales no vamos a decir más que esta breve introducción, para saber más sobre cópulas dinámicas y su aplicación en finanza consúltese [1].



## Aplicación en hidrología

### Introducción

En nuestra siguiente aplicación de las cópulas vamos a ver cómo estas se pueden utilizar para predecir los niveles de afluencia del embalse de Santillana, situado en el río Manzanares, tal y como ha hecho Requena [9]. Para esta investigación se tomaron los datos de 41 años consecutivos en el depósito. Estos datos fueron el caudal máximo anual (Q) y el volumen hidrográfico (V) asociado a dicho caudal.

### Estimación de las marginales

En un estudio previo que se puede encontrar en [5] se mostró que las marginales podían ser modelizadas por distribuciones Gumbel o de valores extremos generalizada. Calculando los parámetros de la Gumbel, elegida entre las dos distribuciones por simplicidad, asociados al caudal máximo y al volumen se obtuvieron los resultados de la Tabla 3.1.

Variable	$\mu$	$\beta$
Q	30,47	22,69
V	5,87	5,70

Tabla 3.1: Valores  $P$  para los ajustes de las cópulas

Estos resultados fueron verificados con el test Kolgomorov-Smirnov obteniendo que dicha hipótesis eran aceptables.

### Estudio de la dependencia

Para ver que forma toma la relación de las variables exponemos la Figura 3.1 de los datos normalizados. Esta expone una clara dependencia positiva, lo que nos da que pensar que las variables pueden no ser independientes. Las medidas de asociación calculadas con las definiciones poblacionales refuerzan esta hipótesis dando  $\tau = 0,722$  y  $\rho = 0,8899$  ambas con una probabilidad  $P < 0,001$ , por lo que se verifica que no se da la independencia.

### Selección de la cópula: estimación y test de ajuste

En este estudio, el abanico de posibilidades considerado para generar un modelo mediante cópulas fue bastante más amplio. Las familias de cópulas consideradas fueron: las cópulas Ali-Mikhail-Haq, Frank, Gumbel, Galambos, Hüsler-Reiss, Tawn, Farlie-Gumbel-Morgenstern y Plackett. Sin embargo, el hecho de que los posibles valores de  $\tau$  para algunas de estas cópulas no contuviesen 0,7244 redujo a 6 las cópulas candidatas. Estas cópulas que no pudieron satisfacer la condición de la Tau de Kendall fueron la Ali-Mikhail-Haq ( $\tau \in [-0,1817, 1/3]$ ), la Tawn ( $\tau \in [0, 0,4184]$ ) y la Farlie-Gumbel-Morgenstern ( $\tau \in [2/9, 2/9]$ ).

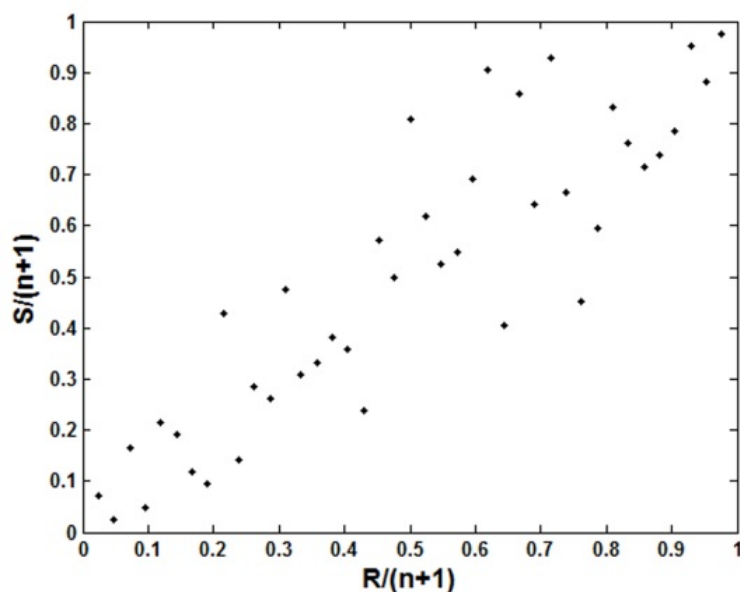


Figura 3.1: Observaciones del caudal máximo y el volumen normalizados por rangos

Para estimar los parámetros de las cópulas restantes se usaron dos métodos: el método de inversión de la Tau de Kendall (ITK) y el método de máxima pseudo-verosimilitud (MPV). Como ya hemos hablado del primero, vamos a explicar en que consiste el método de máxima pseudo-verosimilitud. El método de máxima pseudo-verosimilitud consiste en generar unas pseudo-observaciones, tal y como se hace para el método bootstrap, para las distribuciones marginales, y una vez estas están normalizadas por rangos, hallar el parámetro  $\theta$  que maximiza la función de verosimilitud para la cópula  $C_\theta$  respecto a dichas pseudo-observaciones. En la Tabla 3.2 aparecen los valores de  $\theta$  para cada cópula hallados por los dos métodos descritos.

Familia	$\theta_n$ (según ITK)	$\theta_n$ (según MPV)
Clayton	5,257	3,337
Frank	12,622	11,774
Gumbel	3,628	3,068
Galambos	2,919	2,345
Hüsler-Reiss	3,677	2,970
Plackett	54,230	33,570

Tabla 3.2: Estimaciones del parámetro  $\theta$  para las cópulas según los métodos ITK y MPV

Para decidir que familia de cópulas es la más indicada y que valor del parámetro es el más acertado, Requena halló el estadístico Cramér-von Mises con 10.000 muestras bootstrap  $S_m^*$ . En la Tabla 3.3 se muestran los distintos valores obtenidos para los  $S_n$  y los respectivos valores  $P$  de cara a: comparar los resultados obtenidos con los dos métodos anteriores, descartar las cópulas que menos se ajusten y seleccionar la que mejor lo haga.

Como vemos, el método de la inversión de la Tau de Kendall da en todos los casos mejores resultados que el método de máxima pseudo-verosimilitud, por esto de aquí

Familia	$S_n$ (según ITK)	$S_n$ (según MPV)	$P$ (según ITK)	$P$ (según MPV)
Clayton	0,0223	0,0524	0,3828	0,0565
Frank	0,0174	0,0202	0,8402	0,7332
Gumbel	0,0218	0,0351	0,3967	0,0649
Galambos	0,0219	0,0357	0,3910	0,0603
Hüsler-Reiss	0,0221	0,0379	0,3663	0,0568
Plackett	0,0181	0,0308	0,7893	0,2967

Tabla 3.3: Estadístico  $S_n$  y valor  $P$  para las cópulas con los  $\theta$  hallados con los métodos ITK y MPV

en adelante el estudio se realiza sólo con las cópulas cuyo parámetro ha sido el hallado por el ITK. No obstante, los resultados obtenidos por el otro método también son más que aceptables ya que sus valores  $P$  se matienen por encima de 0,05 y los valores  $S_n$  no son muy altos, esto quiere decir que no los rechazamos, simplemente los desechamos en favor de unos que se adaptan mejor. Por otra parte, no podemos desechar ninguna cópula (como ya hemos seleccionado el candidato de cada familia hablamos de cópulas y no de familias de cópulas) como hicimos en el caso del mercado energético por lo que tenemos que realizar otro análisis que nos permita seleccionar una de las cópulas. En caso de no tener en cuenta otros test, se tomaría la cópula Frank como la que mejor se adapta ya que tiene el  $S_n$  más bajo.

El criterio llevado a cabo para discriminar las cópulas va a ser la dependencia de la cola superior de cada una de las cópulas. En la Tabla 3.4 aparecen tanto los coeficientes de las colas superiores de las cópulas como el coeficiente de la cola superior de la cópula empírica.

Familia	$\lambda_U$
Empirica	0,749
Clayton	0
Frank	0
Gumbel	0,789
Galambos	0,789
Hüsler-Reiss	0,786
Plackett	0

Tabla 3.4: Coeficientes de la cola superior para las cópulas empírica y con el parámetro  $\theta$

Es ahora cuando se eliminan de la lista de candidatas a las cópulas Clayton, Frank y Plackett al no representar apropiadamente el comportamiento cuando se toman valores muy grandes. Esto es bastante relevante de cara a hacer predicciones sobre si se va a rebasar el nivel del embalse o no ya que, en estos casos, se tomarán valores muy altos en ambas variables. De entre las tres restantes elegimos la cópula Gumbel como la que mejor modeliza esta relación por tener un menor  $S_n$ .

Antes de tratar de hacer predicciones con la cópula obtenida vamos a justificar que la cópula elegida necesariamente tiene que representar bien los valores extremos. Para ello, tomamos  $q_T$  y  $v_T$  como los caudales máximos y volúmenes asociados cuando tomamos como periodos de retorno  $T = 10, 100, 1000$ . Una vez tenemos estos valores,

calculamos los tiempos de retorno conjunto: el tiempo de retorno “O” y el tiempo de retorno “Y”. Estos tiempos de retorno son los tiempos que pasan entre una inundación y la siguiente que al menos iguale la magnitud de la primera. Una justificación del uso de estos tiempos de retorno la podemos encontrar en [10]. Para entender mejor esto plasmamos a continuación las fórmulas:

$$T_Q = \frac{1}{\mathbb{P}[Q > q]} = \frac{1}{1 - F(q)}; T_V = \frac{1}{\mathbb{P}[V > v]} = \frac{1}{1 - G(v)}$$

$$T_O = \frac{1}{\mathbb{P}[Q > q \text{ o } V > v]} = \frac{1}{1 - C(F(q), G(v))}$$

$$T_Y = \frac{1}{\mathbb{P}[Q > q \text{ y } V > v]} = \frac{1}{1 - F(q) - G(v) + C(F(q), G(v))}$$

Donde  $F$  y  $G$  son las funciones de distribución del caudal y el volumen respectivamente de las distribuciones Gumbel correspondientes y  $C$  es la cópula seleccionada. Por tanto, dado un tiempo de retorno  $T$  para ambas variables podemos calcular los  $q_T$  y  $v_T$  asociados invirtiendo la primera fórmula y calcular los tiempos de retorno conjuntos. Estos 5 datos aparecen en la Tabla 3.5 para las cópulas Gumbel, la elegida como mejor, y Frank, la que inicialmente parecía mejor, añadiendo además  $t$ , la curva de nivel a la que pertenecen  $q_T$  y  $v_T$ , es decir,  $C(F(q_T), G(v_T)) = t$ .

Cópula	$T = T_Q = T_V$	$q_T$	$v_T$	$t$	$T_O$	$T_Y$
Gumbel	10	81,52	18,69	0,8803	8	12
	100	134,80	32,08	0,9879	83	127
	1000	187,12	45,21	0,9988	826	1266
Frank	10	81,52	18,69	0,8572	7	17
	100	134,80	32,08	0,9811	53	891
	1000	187,12	45,21	0,9980	503	80226

Tabla 3.5: Tiempos de retornos conjuntos simulados para las cópulas Gumbel y Frank

A la vista de los resultados de la Tabla 3.5 es claro que cuando el valor de la cópula  $t$  no es extrema (toma valores muy próximos a 1) el comportamiento los tiempos de retornos conjuntos son similares, sin embargo, cuando este valor se acerca mucho a 1 las diferencias entre las dos cópulas se hacen mucho más notables. Es por esto que era necesario que la cópula se comportase bien en los extremos y por lo que la cópula Gumbel era mejor que la Frank.

### Resultados simulados

Para poder obtener resultados en términos de la capacidad del embalse es necesario generar hidrogramas. Para ello dado una pareja generada  $(Q, V)$  se obtiene su ratio  $Q/V$  y se compara con una base de datos de hidrogramas quedándose con el que tenga el ratio más parecido al generado aleatoriamente. Una vez hecho esto se escala el hidrograma para que el valor máximo sea  $Q$  y después se escala el rango del hidrograma para que el área coincida con  $V$ . Para el siguiente análisis se simularon 100.000 hidrogramas según el proceso comentado. De esta simulación se obtuvo la probabilidad de superar el nivel máximo de agua (NMA) en los hidrogramas anteriores  $p_{exc}$ , luego el

tiempo de retorno de una inundación es puede calcular como  $T_{in} = \frac{1}{p_{exc}}$ . Dados estos datos se puede hallar la curva de nivel de NMA (MWL en inglés) en función del tiempo de retorno de inundación  $T_{in}$  ( $T_{dam}$  e inglés). Dicha curva y algunos de los valores que toma con los datos reales del desagüe de la balsa se encuentran en la Figura 3.2 y la Tabla 3.6. Estas medidas son: una altura de 889 m y una longitud de 12 m.

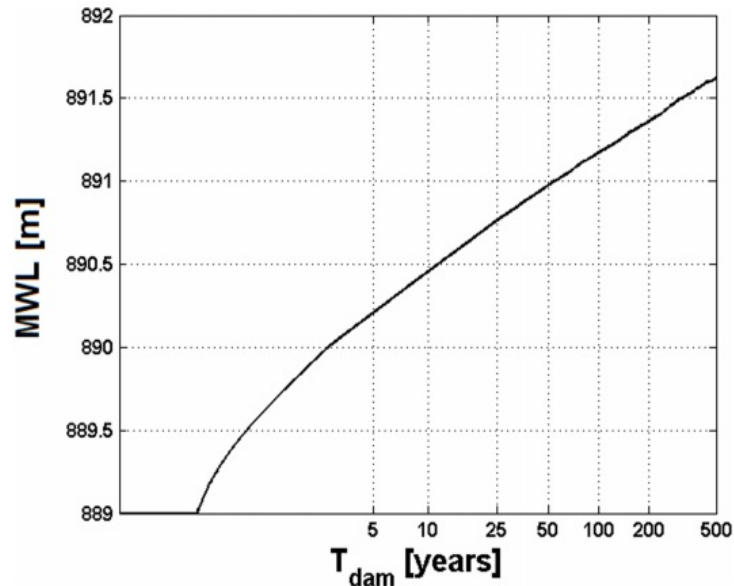


Figura 3.2: Curva de frecuencia del NMA para las medidas reales del desagüe

$T_{in}$ [años]	NMA[m]
5	890,21
10	890,46
50	890,98
100	891,17
500	891,62

Tabla 3.6: Valores obtenidos de Figura 3.2

Eligiendo los hidrogramas que presentan un mismo NMA podemos generar curvas en el espacio del caudal y el volumen para los tiempos de retorno de inundación deseados. Esto es, dado un tiempo de retorno de inundación calculamos el NMA asociado y tomamos de nuestra simulación todos los hidrogramas con el mismo NMA y representamos estos hidrogramas como las parejas ( $Q, V$ ) de las que se crearon. Estas curvas de nivel junto con las curvas de nivel de los tiempos de retorno  $T_O$  y  $T_Y$  están representadas en la Figura 3.3 con el fin de compararlas.

Tenemos que resaltar de la Figura 3.3 que las curvas del tiempo de retorno “O” se acercan mucho más a las curvas del tiempo de retorno de inundación. Esto es útil ya que para predecir inundaciones se pueden usar las curvas del tiempo de retorno “O” en lugar de las de inundación ya que estas últimas son más difíciles de calcular y más imprecisas por como se han generado. En el texto [9] el análisis llega más allá

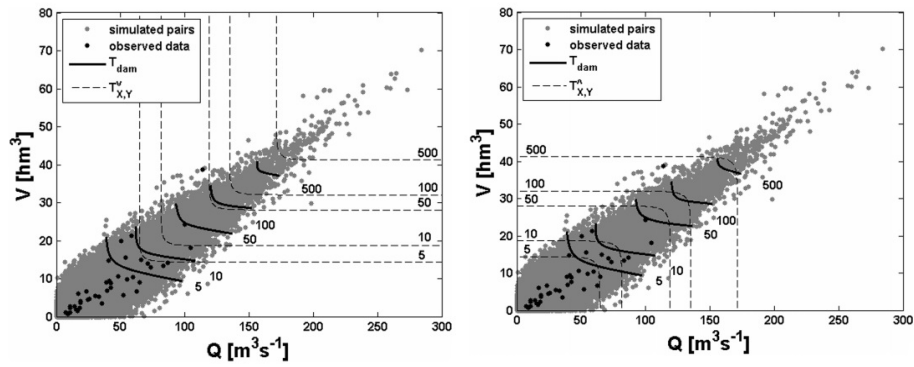


Figura 3.3: Comparación entre los tiempos de retorno de inundación y los tiempos de retorno condicionados “O”(a la izquierda) e “Y”(a la derecha)

considerando también el tiempo de retorno de Kendall, que se basa en las curvas de nivel de la cópula y no en las probabilidades “O/Y” de la misma, obteniendo que dichas curvas se asemejan mejor a las de inundación que las de “O” y por tanto son una mejor estimación.

A modo de ejemplo, generamos aleatoriamente una pareja de valores extremos y los tiempos de retorno que le corresponderían, para ello seleccionamos el valor de la cópula en dichos puntos. Los valores  $t$  de la cópula van a ser 0,9 y 0,99, de ellos podemos simular una pareja  $(Q, V)$  como hemos hecho hasta ahora simplemente tomando un valor de la uniforme  $U[0, 1]$ . En la Tabla 3.7 recogemos la información que hasta ahora hemos ido obteniendo para una pareja de valores para los  $t$  mencionados.

$t$	$Q$	$V$	$T_O$	$T_Q$	$T_V$	$T_Y$	NMA[m]	$T_{in}$
0,9	90,52	19,12	10	15	11	16	890,53	13
0,99	136,41	34,43	100	107	151	168	891,34	180

Tabla 3.7: Simulación con valores extremos con  $t = 0,9, 0,99$ .

Finalmente, vamos a suponer que la altura del desagüe  $E$  toma también los valores de 884 m y 879 m además del real (889 m) y que la longitud  $L$  también varía, con valores de 7 m, 12 m (el real) y 17 m. Con objeto de comparar las distintas curvas que nivel que toma  $T_{in}$  para los años de la tabla anterior, las presentamos en la Figura 3.4.

Lo que queremos resaltar de este grupo de gráficas es como cambia la curva del tiempo de retorno de inundación conforme modificamos los parámetros del desagüe. Cuanto menor es la altura del desagüe más vertical es la curva, por tanto, las inundaciones dependerán más del caudal máximo registrado en vez del volumen del mismo. Equivalentemente, cuanto mayor es la altura, más fuerza tomará el volumen de la creciente a la hora de predecir si habrá o no una inundación. Justo lo contrario pasa con la anchura del desagüe: cuanto mayor sea más relevante será el caudal ya que la curva será más vertical y cuanto menor sea la anchura mayor será la importancia del volumen.

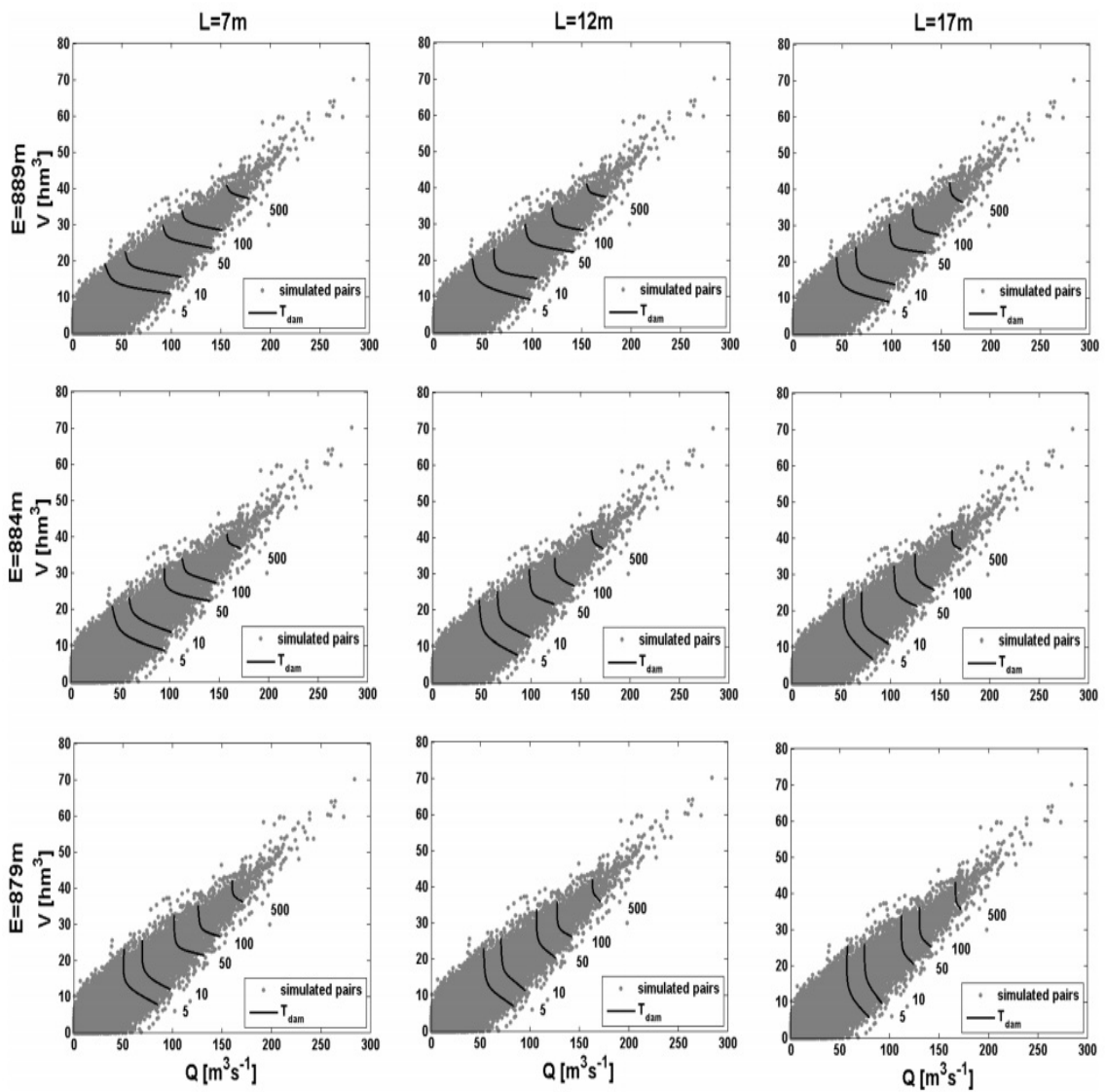


Figura 3.4: Comparación de las curvas de nivel de inundación y las medidas del desagüe del embalse.





## Conclusiones

En 1951, Fréchet planteó el problema de cómo expresar la dependencia entre las de variables aleatorias que componen un vector aleatorio. Fue Sklar, en 1959, quien dio solución a este problema introduciendo el concepto de cópula. Por medio del Teorema de Sklar se consigue expresar la relación que existe entre las funciones de distribución marginales, la función de distribución conjunta de un vector aleatorio y el concepto de cópula. Es por esto que las cópulas conforman una herramienta conceptualmente asequible al ser, desde el punto de vista probabilístico, la expresión funcional de la dependencia. Sin embargo, han empezado a cobrar un espectacular interés en el mundo de la investigación a partir de su aplicación en el mundo de las finanzas. Desde entonces, el abanico de posibilidades que ofrecen las cópulas ha seguido creciendo abarcando áreas como las ciencias experimentales, en especial la hidrología; las ciencias actuariales, para calcular mortalidad de cara a seguros de vida; la biomedicina, para modelar los tiempos de infección en órganos pares; o la ingeniería, para procesos multivariados. Es por todo esto que el estudio de las cópulas resulta de interés, ofreciendo herramientas sencillas que se pueden aplicar en un vasto elenco de campos.

Esta memoria introduce los tópicos básicos para acceder al conocimiento de las cópulas y dos estudios de aplicaciones de las mismas; uno en hidrología y otro en finanzas.



## Bibliografía

- [1] U. Cherubini, F. Gobbi, S. Mulinacci, S. Romagnoli, *Dynamic copula methods in finance*, Wiley Finance, 2012.
- [2] G. Frahm, M. Junker, R. Schmidt *Estimating the tail-dependence coefficient: Properties and pitfalls*, Insurance, Mathematics and Economy, **37**, 2005, 80–100.
- [3] C. Genest, B. Rémillard, *Validity of the parametric bootstrap for goodness-of-fit testing in semiparametric models*, Annales de l’institut Henri Poincaré, **6** (2008), 1096–1127.
- [4] V. Grégoire, C. Genest, M. Gendrou, *Using copulas to model price dependence in energy markets*, Energy risk, **Marzo** 2008, 62–68.
- [5] A. Jiménez Álvarez, C. García Montañés, L. Mediero Orduña, L. Incio Caballer, J. Garrote Revilla, *El mapa de caudales máximos de las cuencas intercomunitarias*, Revista de obras públicas, **3533** (2012), 7–32.
- [6] D.C. Moreno Chavarro *Método para elegir una cópula arquimediana óptima*, Universidad Nacional de Colombia, 2012.
- [7] R.G. Nelsen, *An introduction to copulas*, Springer, 2006 (2nd Ed.).
- [8] Quantdare                       $\text{\TeX}$ :                      <http://quantdare.com/2014/03/por-que-usar-rendimientos-logaritmicos/>.
- [9] A.I. Requena, L. Mediero, L. Garrote, *A bivariate return period based on copulas for hydrologic dam desing: accounting for reservoir routing in risk estimation*, Hydrology and Earth system sciences, **17**, 2013, 3023–3038.
- [10] J.T. Shiau, H.Y. Wang, C.T. Tsai, *Bivariate frequency analysis of flooding using copulas*, Journal of the American water resources association, **6** (2006), 1549–1564.