



Facultad de Ingeniería  
Universidad de la República



Tesis de Maestría en Ingeniería Matemática  
Director: Ernesto Mordecki

---

# Cóputas paramétricas y no paramétricas

Con aplicaciones en riesgo bancario

---

Gabriel Illanes

FEBRERO 2013  
MONTEVIDEO - URUGUAY

# Índice general

<b>Introducción</b>	<b>3</b>
<b>1. Dependencia entre variables aleatorias, cópulas</b>	<b>6</b>
1.1. Definición y propiedades de una cópula . . . . .	7
1.2. Teorema de Sklar. . . . .	11
1.3. Vectores aleatorios y cópulas. . . . .	15
1.4. Medidas clásicas de dependencia . . . . .	18
1.4.1. Tau de Kendall . . . . .	19
1.4.2. Rho de Spearman . . . . .	20
1.5. Dependencias de colas . . . . .	22
<b>2. Estimación no paramétrica de cópulas</b>	<b>25</b>
2.1. Cópula empírica . . . . .	26
2.2. Estimación de densidad por random forests . . . . .	27
2.3. Convergencia . . . . .	31
2.4. Aplicaciones numéricas . . . . .	36
<b>3. Estimación paramétrica</b>	<b>40</b>
3.1. Familias arquimedianas . . . . .	41
3.2. Estimación de parámetros por máxima verosimilitud. . . . .	46
3.3. Bondad de ajuste, test chi cuadrado . . . . .	48
<b>4. Aplicación a datos bancarios</b>	<b>50</b>
4.1. Datos y modelo. . . . .	50
4.1.1. Fórmula de Black Scholes . . . . .	50
4.1.2. Modelo estructural . . . . .	51
4.1.3. Credit Default Swaps . . . . .	52
4.1.4. Nuestros datos . . . . .	54
4.2. Resultados: caso no paramétrico . . . . .	57
4.3. Resultados: caso paramétrico . . . . .	58

4.4. Indicadores de riesgo bancario . . . . .	63
<b>Conclusiones</b>	<b>65</b>
<b>Apéndice</b>	<b>67</b>

# Introducción

En los modelos estadísticos, los fenómenos con comportamiento incierto son modelados a través de variables aleatorias. Cuando observamos dos variables aleatorias  $X_1$  y  $X_2$ , o bien sucede que los valores de  $X_1$  y  $X_2$  no guarden ninguna relación entre sí, o bien existe algún tipo de dependencia. Ejemplos de dependencia son  $X_1 = X_2$ , o que  $(X_1, X_2)$  formen un vector aleatorio normal con correlación  $\rho$ . Muchas veces se intenta describir la dependencia entre dos variables aleatorias con la correlación lineal, pero esto tiene varias desventajas. Algunas de ellas son que la correlación lineal no es invariante bajo transformaciones estrictamente crecientes de las variables aleatorias; es posible encontrar variables aleatorias que no admitan correlaciones lineales iguales a 1 o -1; y correlación lineal igual a cero no implica independencia, más aún, no podemos asegurar que no se cumple  $Y = \alpha(X)$  para alguna función estrictamente creciente  $\alpha$  (por más detalles ver [10] sección 3.1.4).

Sin embargo, es posible rescatar toda la información de la dependencia entre dos variables aleatorias con funciones que cumplen ciertas características. Esta tesis apunta a estudiar dichas funciones, a las cuales llamaremos *cóputas*.

Por más que el concepto y nombre de *cóputa* surgió con el paper de Sklar ([13]) en 1959 (incluso el concepto de *cóputa* bajo otro nombre ya había aparecido antes en la literatura, por ejemplo en [14], 1940), no es hasta los años 90 que empieza el estudio sistemático de la teoría de *cóputas* y sus aplicaciones.

Las *cóputas* pueden pensarse de dos maneras. Una manera es pensar las *cóputas* como la función que junta o enlaza (el significado de la palabra “*cóputa*” en latín)  $n$  funciones de distribución unidimensionales  $F_i$ , para generar una función de distribución conjunta  $n$ -dimensional con marginales  $F_i$ . La otra manera es considerar a las *cóputas* como una función de distribución conjunta  $n$ -dimensional con marginales uniformes  $U(0, 1)$ . En este traba-

jo nos concentraremos únicamente en el caso  $n = 2$ . A través del teorema de Sklar es que interpretamos las cópulas como las funciones que rescatan la estructura de dependencia de una función de distribución conjunta. La teoría de cópulas, en particular, ayuda a la comprensión y estudio de la dependencia en eventos extremos; dentro de esto estudiaremos el concepto de *dependencia de colas*. Además, las cópulas son una herramienta muy potente para generar funciones de distribución multidimensionales dadas las marginales.

En el capítulo 1 nos dedicaremos a presentar formalmente el concepto de cópula, sus propiedades básicas, y ahondaremos en sus distintas interpretaciones y utilidades.

En nuestro caso, pretendemos utilizar las herramientas que provee la teoría de cópulas al área de la economía. Nuestro objetivo será estudiar la dependencia existente entre los valores de activos de 5 bancos de interés: Citibank NA, Banco Santander Central Hispano SA, BBVA SA, HSBC Bank USA y Credit Lyonnais. Sobre todo, nos interesará estudiar la dependencia en momentos de crisis: dado que uno de estos bancos tiene bajo rendimiento, ¿cómo afectará eso al desempeño del resto de los bancos?. Enmarcamos este trabajo como continuación a la tesis de Maestría en Economía Internacional de Verónica Rodríguez, “Indicadores de Riesgo derivados de Depósitos Bancarios Constituidos en el Exterior” ([9]).

Este problema es complicado por varias razones, pero sobre todo a la falta de información: no tenemos acceso al valor de los activos de los bancos. Sin embargo sí contamos con información sobre los valores de *Credit Default Swaps* (o CDS). Los CDS pueden pensarse como un seguro al que puede acceder un prestamista para protegerse contra el evento en donde el prestatario entra en default; sin embargo, lo particular de los CDS es que cualquier persona puede comprar este seguro, no solo el prestamista (en cuyo caso están apostando al default del prestatario para conseguir ganancia). Estos precios de CDS llevan implícito la probabilidad de default del prestatario, y a través de ellos podemos rescatar la dependencia que buscamos (la metodología se basará en [9], nosotros aportaremos la teoría de cópulas para mejorarlo).

Dedicaremos el capítulo 4 a desarrollar la teoría en la que se enmarcan los CDS y a obtener los resultados deseados.

Teniendo en cuenta nuestro objetivo, necesitamos herramientas para poder estimar las cópulas implícitas en un conjunto de observaciones (o sea, dada una muestra aleatoria simple, estimar la cópula de la cual provienen).

De manera similar a la estimación de distribuciones conjuntas (y de sus densidades), tenemos métodos paramétricos y no paramétricos.

Para la estimación no paramétrica propondremos un estimador de densidades de cópulas basado en el método conocido como “random forests” (ver [2]), y estudiaremos algunas de sus propiedades. Estas estimaciones tienen la gran ventaja que no tenemos que imponer ninguna suposición sobre la cópula de la cual provienen nuestros datos, y se puede extraer información útil para conjeturar acerca de la verdadera cópula (por ejemplo, en dónde se dan los máximos y mínimos de la densidad), pero se pierde precisión cuando los valores de la densidad son muy altos. Ampliaremos esto en el capítulo 2.

Para la estimación paramétrica estudiaremos las familias de cópulas *arquimedianas*, veremos algunas de sus propiedades, y utilizaremos 4 familias como candidatas para la estimación paramétrica: estas familias son la Gumbel, Frank, Clayton, y la biparamétrica. Para el ajuste utilizaremos el método de máxima verosimilitud, y como método de bondad de ajuste utilizaremos un test chi cuadrado. La ventaja de estos métodos es notoria en el caso de obtener buenos valores de ajuste: tendremos una cópula con una expresión manejable a la cual le podremos aplicar toda la teoría de cópulas arquimedianas. Además, incluso sin tener buenos valores de ajuste, las estimaciones paramétricas pueden aportar información sobre la dependencia de colas.

Este trabajo fue propuesto por Alejandro Pena (Jefe de Departamento Riesgos Financieros del Banco Central del Uruguay), y se enmarca como la continuación de la tesis de Maestría en Economía Internacional de Verónica Rodríguez, “Indicadores de Riesgo derivados de Depósitos Bancarios Constituidos en el Exterior” ([9]).

# Capítulo 1

## Dependencia entre variables aleatorias, cópulas

A continuación, daremos una introducción a las cópulas, basándonos en el capítulo 2 y 5 de [1].

Para fijar notación, consideremos dos variables aleatorias  $X, Y$ , con funciones de distribución  $F_X, F_Y: \mathbb{R} \rightarrow \mathbf{I}$ , donde  $\mathbf{I} = [0, 1]$  (se tendrá que  $F_X(a) = \mathbb{P}\{X \leq a\}$  y  $F_Y(b) = \mathbb{P}\{Y \leq b\}$ ), y una función de distribución conjunta  $F_{XY}: \mathbb{R}^2 \rightarrow \mathbf{I}$  (o sea, la distribución del vector aleatorio  $(X, Y)$ ,  $F_{XY}(a, b) = \mathbb{P}\{X \leq a, Y \leq b\}$ ). Por comodidad, a veces pensaremos que dichas distribuciones están definidas en la recta ampliada, o sea,  $F_X: \bar{\mathbb{R}} \rightarrow \mathbf{I}$ , donde  $\bar{\mathbb{R}} = [-\infty, +\infty]$ , y  $F_{XY}: \bar{\mathbb{R}}^2 \rightarrow \mathbf{I}$ . De existir, denotaremos por  $f_X = F'_X = \frac{\partial}{\partial x} F_X$  a la densidad de  $X$  (análogamente definimos  $f_Y = \frac{\partial}{\partial y} F_Y$ ), y por  $f_{XY} = \frac{\partial^2}{\partial x \partial y} F_{XY}$  a la densidad del vector aleatorio  $(X, Y)$ .

Conociendo la distribución conjunta  $F_{XY}$ , podemos calcular la distribución de las marginales como  $F_X(a) = F_{XY}(a, +\infty)$ ,  $F_Y(b) = F_{XY}(+\infty, b)$ . De forma inversa, si solo conociéramos las marginales  $F_X$  y  $F_Y$ , y si supiéramos que  $X$  e  $Y$  son independientes, podríamos afirmar que la distribución conjunta es  $F_{XY}(a, b) = F_X(a)F_Y(b)$ . Sin embargo, en general, no podemos calcular la distribución conjunta  $F_{XY}$  conociendo únicamente las distribuciones  $F_X$  y  $F_Y$ . En este capítulo mostraremos como, conociendo  $F_{XY}$ , es posible encontrar una función  $C_{XY}: \mathbf{I}^2 \rightarrow \mathbf{I}$  tal que  $F_{XY}(a, b) = C_{XY}(F_X(a), F_Y(b))$  para todo  $(a, b) \in \mathbb{R}^2$  (que además será única si  $F_X$  y  $F_Y$  son continuas), a la que llamaremos **cópula del vector aleatorio**  $(X, Y)$  (primera sección), luego enunciaremos y demostraremos el teo-

rema de Sklar, el cual nos permite entender el papel que juegan las cópulas en el vínculo entre una función de distribución conjunta y sus marginales (segunda sección), proseguiremos interpretando las cópulas como la función que expresa completamente la *dependencia* entre dos variables aleatorias  $X$  e  $Y$  (tercera sección), estudiaremos algunas maneras de medir concordancia usando cópulas (cuarta sección), para finalizar con el estudio de dependencia en eventos extremos (quinta sección).

## 1.1. Definición y propiedades de una cópula

Nuestro primer objetivo es introducir la noción de funciones bivariadas **2-crecientes**, las cuales serán un análogo a las funciones no decrecientes de una variable. Para ello, necesitamos introducir más notación. Llamaremos **rectángulo** en  $\overline{\mathbb{R}}^2$  al producto  $[x_1, x_2] \times [y_1, y_2]$  de dos intervalos en  $\overline{\mathbb{R}}$ ; los **vértices** de dicho rectángulo serán  $(x_1, y_1)$ ,  $(x_1, y_2)$ ,  $(x_2, y_1)$  y  $(x_2, y_2)$ . Una **función bivariada real** será una función  $h$  cuyo dominio ( $\text{Dom}h$ ) sea un subconjunto de  $\overline{\mathbb{R}}^2$  y su rango ( $\text{Ran}h$ ) sea un subconjunto de  $\overline{\mathbb{R}}$ .

**Definición 1.1.1.** Sean  $S_1$  y  $S_2$  subconjuntos no vacíos de  $\overline{\mathbb{R}}$ , y sea  $h$  una función bivariada real, tal que  $\text{Dom}h = S_1 \times S_2$ . Sea  $B = [x_1, x_2] \times [y_1, y_2]$  un rectángulo tal que sus vértices estén contenidos en  $\text{Dom}h$ . Definimos el  **$h$ -volumen del rectángulo  $B$**  como

$$V_h(B) = h(x_2, y_2) - h(x_2, y_1) - h(x_1, y_2) + h(x_1, y_1).$$

**Definición 1.1.2.** Sea  $h$  una función bivariada real tal que  $\text{Dom}h = S_1 \times S_2$ , con  $S_1, S_2 \subset \overline{\mathbb{R}}$ . Decimos que  $h$  es **2-creciente** si  $V_h(B) \geq 0$  para todo rectángulo  $B$  cuyos vértices estén contenidos en  $\text{Dom}h$ .

Otra manera de pensar a las funciones 2-crecientes es la siguiente:

**Lema 1.1.3.** Sea  $h$  una función bivariada real tal que  $\text{Dom}h = S_1 \times S_2$ , con  $S_1, S_2 \subset \overline{\mathbb{R}}$ . Entonces  $h$  es 2-creciente si y solo si, cualesquiera sean  $x_1, x_2 \in S_1$ ,  $y_1, y_2 \in S_2$  tal que  $x_1 \leq x_2$ ,  $y_1 \leq y_2$ , los mapas  $t \xrightarrow{f} h(x_2, t) - h(x_1, t)$  y  $s \xrightarrow{g} h(s, y_2) - h(s, y_1)$  son no decrecientes.

La demostración de este lema es inmediata, y surge de notar que si  $t_1, t_2 \in S_1$  y  $t_1 \leq t_2$ , entonces  $f(t_2) - f(t_1) = V_h(B)$  para el rectángulo  $B$  de vértices  $(t_1, y_1)$ ,  $(t_1, y_2)$ ,  $(t_2, y_1)$  y  $(t_2, y_2)$  (idem con el mapa  $g$ ).

Vale observar que una función  $h$  2-creciente no necesariamente es no decreciente en cada una de sus variables (o sea, no necesariamente cumple que  $h(\cdot, y_0)$  y  $h(x_0, \cdot)$  sean no decrecientes).



**Ejemplo.** Sea  $h: \mathbf{I}^2 \rightarrow \mathbb{R}$  definida por  $h(x, y) = (2x - 1)(2y - 1)$ , y sea  $B = [x_1, x_2] \times [y_1, y_2] \subset \mathbf{I}^2$ . Entonces

$$\begin{aligned} V_h(B) &= (2x_2 - 1)(2y_2 - 1) - (2x_2 - 1)(2y_1 - 1) \\ &\quad - (2x_1 - 1)(2y_2 - 1) + (2x_1 - 1)(2y_1 - 1) \\ &= (2x_2 - 1)(2y_2 - 2y_1) - (2x_1 - 1)(2y_2 - 2y_1) \\ &= (2x_2 - 2x_1)(2y_2 - 2y_1) \geq 0 \end{aligned}$$

y por lo tanto  $h$  es 2-creciente. Sin embargo, fijado  $y_0 \in (0, \frac{1}{2})$ , se tiene que  $h(\cdot, y_0)$  es decreciente (análogamente, fijado  $x_0 \in (0, \frac{1}{2})$ , se tiene que  $h(x_0, \cdot)$  es decreciente).

Tampoco podemos afirmar que  $h$  sea 2-creciente partiendo que  $h$  es no decreciente en cada variable.

**Ejemplo.** Sea  $h: \mathbf{I}^2 \rightarrow \mathbb{R}$  definida por  $h(x, y) = \max(x, y)$ . Claramente,  $h(\cdot, y_0)$  y  $h(x_0, \cdot)$  son no decrecientes  $\forall x_0, y_0 \in \mathbf{I}$ . Sin embargo  $V_h(\mathbf{I}^2) = -1$ , por lo tanto  $h$  no es 2-creciente.

Sin embargo, sabiendo que  $h$  es 2-creciente y con una hipótesis extra (pediremos que  $h$  sea basada), podemos asegurarnos que  $h$  es no decreciente en cada variable.

**Definición 1.1.4.** Sea  $h: S_1 \times S_2 \rightarrow \overline{\mathbb{R}}$  una función 2-creciente. Supongamos que existe  $m_1$  el mínimo de  $S_1$  (eventualmente  $m_1 = -\infty$ ) y supongamos que existe  $m_2$  el mínimo de  $S_2$ . Decimos que  $h$  está **basada** si  $h(m_1, y) = 0 = h(x, m_2) \forall (x, y) \in S_1 \times S_2$ .

**Lema 1.1.5.** Sea  $h: S_1 \times S_2 \rightarrow \overline{\mathbb{R}}$  una función 2-creciente y basada. Entonces  $h$  es no decreciente en cada variable.

*Demostración.* Demostraremos que  $h$  es no decreciente fijando la segunda coordenada, es análogo para la primera coordenada.

Sea  $m_2$  el mínimo de  $S_2$ , y fijemos  $y_0 \in S_2$ . Combinando el lema 1.1.3 con que  $h$  está basada tenemos que  $t \xrightarrow{f} h(t, y_0) - h(t, m_2) = h(t, y_0)$  es no decreciente, con lo cual obtenemos lo deseado.  $\square$

Supongamos ahora que  $S_1$  tiene un máximo, al cual llamaremos  $M_1$  (eventualmente  $M_1 = +\infty$ ), y supongamos que existe  $M_2$  el máximo de  $S_2$ . Decimos que  $h: S_1 \times S_2 \rightarrow \overline{\mathbb{R}}$  tiene **marginales**, las cuales son funciones  $i: S_1 \rightarrow \overline{\mathbb{R}}$  y  $j: S_2 \rightarrow \overline{\mathbb{R}}$ , definidas por

$$\begin{aligned} i(a) &= h(a, M_2) \\ j(b) &= h(M_1, b) \end{aligned}$$

Enunciamos a continuación un lema que nos será útil.

**Lema 1.1.6.** *Sea  $h: S_1 \times S_2 \rightarrow \mathbb{R}$  una función 2-creciente y basada, y sean  $i, j$  las marginales de  $h$ . Sean  $(x_1, y_1), (x_2, y_2) \in S_1 \times S_2$ , entonces*

$$|h(x_2, y_2) - h(x_1, y_1)| \leq |i(x_2) - i(x_1)| + |j(y_2) - j(y_1)|$$

*Demostración.* Utilizando la desigualdad triangular, obtenemos

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

Supondremos sin pérdida de generalidad que  $x_1 \leq x_2$ , y que  $y_1 \leq y_2$ . Sabemos entonces, que como  $h$  es 2-creciente y basada, es creciente en cada una de sus variables, por lo tanto  $h(x_2, y_2) - h(x_1, y_2) \geq 0$ . Utilizando además el lema 1.1.3 obtenemos

$$0 \leq h(x_2, y_2) - h(x_1, y_2) \leq h(x_2, M_2) - h(x_1, M_2) = i(x_2) - i(x_1)$$

Intercambiando los roles de  $x_2$  y  $x_1$  obtenemos  $|h(x_2, y_2) - h(x_1, y_2)| \leq |i(x_2) - i(x_1)|$ , y análogamente obtenemos  $|h(x_1, y_2) - h(x_1, y_1)| \leq |j(y_2) - j(y_1)|$ .  $\square$

Estamos en condiciones de definir las cópulas. Lo haremos en función de las propiedades básicas que cumplen, y más adelante interpretaremos su relación con la dependencia entre variables aleatorias. La siguiente definición es más general que la de cópulas, y nos permite trabajar con vectores aleatorios con funciones de distribución marginales discontinuas.

**Definición 1.1.7.** Una **subcópula bidimensional** (o simplemente **subcópula** si se sobrentiende la dimensión) es una función  $C'$  con las siguientes propiedades:

1.  $\text{Dom}C' = S_1 \times S_2$ , donde  $S_1$  y  $S_2$  son subconjuntos de  $\mathbf{I}$  que contienen al 0 y al 1.
2.  $C'$  es 2-creciente y basada.
3. Para todo  $(u, v) \in S_1 \times S_2$  se tiene  $C'(u, 1) = u$ ,  $C'(1, v) = v$ .

**Definición 1.1.8.** Una **cópula bidimensional** (o simplemente **cópula**) es una subcópula cuyo dominio es  $\mathbf{I}^2$ .

Enunciaremos a continuación algunas propiedades de las cópulas.

*Observación 1.1.9.* Sea  $C$  una cópula y  $S_1 \times S_2$  el dominio de una subcópula. Entonces  $C|_{S_1 \times S_2}$  es una subcópula.

**Teorema 1.1.10.** *Cotas de Fréchet-Hoeffding*

Sea  $C'$  una subcópula. Entonces, para todo  $(u, v) \in S_1 \times S_2$  se tiene

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

*Demostración.* Sea  $(u, v) \in S_1 \times S_2$ , entonces  $C'(u, v) \leq C'(u, 1) = u$ , donde la desigualdad se da por ser  $C'$  una función 2-creciente y basada (y por el lema 1.1.5 es no decreciente en sus coordenadas). De manera análoga tenemos  $C'(u, v) \leq v$ , y por tanto  $C'(u, v) \leq \min(u, v)$ .

Además, si llamamos  $B$  al rectángulo de vértices  $(1, 1)$ ,  $(u, 1)$ ,  $(1, v)$  y  $(u, v)$ , tenemos que  $V_{C'}(B) = C'(1, 1) - C'(u, 1) - C'(1, v) + C'(u, v) \geq 0$ , y por lo tanto  $C'(u, v) \geq u + v - 1$ . Como también se cumple  $C'(u, v) \geq 0$ , obtenemos la otra desigualdad.  $\square$

No solo podemos encontrar una cota para una (sub)cópula cualquiera, sino que tanto  $W$  como  $M$  son cópulas, por lo cual las cotas de *Fréchet-Hoeffding* son las mejores cotas que podemos pedir.

*Observación 1.1.11.*  $W$  y  $M$  son cópulas.

*Demostración.* Es inmediato verificar todas las propiedades de cópulas de  $W$  y  $M$ , salvo que sean 2-crecientes. Sea  $B$  un rectángulo de vértices  $(u_1, v_1)$ ,  $(u_1, v_2)$ ,  $(u_2, v_1)$  y  $(u_2, v_2)$ . Entonces, para demostrar que  $M$  es 2-creciente, supongamos  $u_1 \leq v_1$  sin pérdida de generalidad, y tenemos

$$\begin{aligned} V_M(B) &= M(u_2, v_2) - M(u_2, v_1) - M(u_1, v_2) + M(u_1, v_1) \\ &= M(u_2, v_2) - M(u_2, v_1) \geq 0 \end{aligned}$$

Para demostrar que  $W$  es 2-creciente, observemos primero que  $f(x, y) = x + y - 1$  es creciente en ambas coordenadas. Por tanto, si  $W(u_2, v_2) = 0$ , entonces  $W(u_i, v_j) = 0$  con  $i, j = 1, 2$ . De manera análoga, si  $W(u_1, v_2) = 0$  o si  $W(u_2, v_1) = 0$  entonces  $W(u_1, v_1) = 0$ . Supongamos entonces que  $W(u_i, v_j) > 0$  para todo  $i, j$  con  $ij \neq 1$ , entonces

$$\begin{aligned} W(u_2, v_1) + W(u_1, v_2) &= u_2 + v_2 - 1 + (u_1 + v_1 - 1) \\ &\geq W(u_2, v_2) + W(u_1, v_1) \end{aligned}$$

$\square$

Además, tenemos que las cópulas son funciones lipschitzianas, y por lo tanto continuas (esto surge inmediatamente del lema 1.1.6):

**Teorema 1.1.12.** Sea  $C'$  una subcópula, entonces  $|C'(u_2, v_2) - C'(u_1, v_1)| \leq |u_2 - u_1| + |v_2 - v_1|$ .

## 1.2. Teorema de Sklar.

A partir de ahora, estudiaremos las cópulas en el contexto de *funciones de distribución* de vectores aleatorios (unidimensionales y bidimensionales). Por comodidad, definiremos a las funciones de distribución en la recta (plano) real extendida.

**Definición 1.2.1.** Una **función de distribución (unidimensional)**  $F$  es una función definida en  $\overline{\mathbb{R}}$ , no decreciente, que cumple, que cumple  $F(-\infty) = 0$  y  $F(+\infty) = 1$ .

**Definición 1.2.2.** Una **función de distribución conjunta**  $H$  es una función definida en  $\overline{\mathbb{R}}^2$ , 2-creciente, basada, que cumple  $H(+\infty, +\infty) = 1$ .

*Observación 1.2.3.* Sea  $H$  una función de distribución conjunta, y sean  $F(x) = H(x, +\infty)$  y  $G(y) = H(+\infty, y)$  sus marginales. Entonces, por el lema 1.1.6, se tiene que  $F$  y  $G$  son funciones de distribución.

Notemos que, dada una variable aleatoria  $X$ , su función de distribución  $F_X$  cumplirá la definición de función de distribución (unidimensional) recién enunciada. Sin embargo, en nuestra teoría, no partimos de variables aleatorias, solo importan las funciones de distribución.

Enunciaremos a continuación el teorema de Sklar (1959), lo demostraremos, y luego estudiaremos sus interpretaciones e implicaciones.

**Teorema 1.2.4.** *Teorema de Sklar*

Sea  $H$  una distribución conjunta con marginales  $F$  y  $G$ . Entonces existe una cópula  $C$  tal que, para todo  $x, y \in \overline{\mathbb{R}}$ ,

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

Si  $F$  y  $G$  son continuas, entonces  $C$  es única. De lo contrario,  $C$  está unívocamente determinada en  $\text{Ran}F \times \text{Ran}G$ . Recíprocamente, si  $F$  y  $G$  son funciones de distribución, y  $C$  es una cópula, entonces  $H(x, y) := C(F(x), G(y))$  es una función de distribución conjunta con marginales  $F$  y  $G$ .

La palabra *cópula* fue elegida a razón que la función  $C$  *acopla* las marginales  $F$  y  $G$  para recuperar la distribución conjunta  $H$ . Además, dada una distribución conjunta  $H$ , podemos obtener a partir de ella sus dos marginales  $F$ ,  $G$ , y la cópula  $C$  del teorema de Sklar.

Para demostrar el teorema de Sklar necesitamos dos lemas previos.

**Lema 1.2.5.** *Sea  $H$  una función de distribución conjunta  $H$  con marginales  $F$  y  $G$ . Entonces existe una única subcópula  $C'$  tal que*

- $\text{Dom}C' = \text{Ran}F \times \text{Ran}G$ .
- Para todo  $x, y \in \overline{\mathbb{R}}$ , se tiene  $H(x, y) = C'(F(x), G(y))$ .

*Demostración.* La distribución conjunta  $H$  entra en las hipótesis del lema 1.1.6 donde  $S_1 = S_2 = \overline{\mathbb{R}}$ , por lo tanto

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

de lo cual deducimos que si  $F(x_2) = F(x_1)$  y si  $G(y_2) = G(y_1)$  se tiene que  $H(x_2, y_2) = H(x_1, y_1)$ . Entonces los pares

$$\left\{ \left( (F(x), G(y)), H(x, y) \right) \mid x, y \in \overline{\mathbb{R}} \right\}$$

inducen una función bivariada  $C'$  cuyo dominio es  $\text{Ran}F \times \text{Ran}G$ . Basta verificar que  $C' : \text{Ran}F \times \text{Ran}G \rightarrow [0, 1]$  es una subcópula. Como  $F(-\infty) = G(-\infty) = 0$  y  $F(+\infty) = G(+\infty) = 1$ , tenemos que  $\{0, 1\} \subset \text{Dom}C'$ . Además, si  $u \in \text{Ran}F$  entonces existe  $x \in \overline{\mathbb{R}}$  tal que  $F(x) = u$ , y  $C'(u, 0) = H(F(x), G(-\infty)) = 0$  ya que  $H$  es basada, y por tanto (con un argumento análogo para la otra variable)  $C'$  es basada. De la misma manera,  $C'$  hereda la propiedad de ser 2-creciente de  $H$ . Por último, si  $u$  es como antes,  $C'(u, 1) = H(x, +\infty) = F(x) = u$  (análogo para la otra variable).  $\square$

**Lema 1.2.6.** *Sea  $C'$  una subcópula, entonces existe una cópula  $C$  tal que  $C(u, v) = C'(u, v) \forall u, v \in \text{Dom}C'$ . En otras palabras, toda subcópula se puede extender (no necesariamente de manera única) a una cópula.*

*Demostración.* Sea  $S_1 \times S_2$  el dominio de  $C'$ . Denotemos por  $\overline{S}$  a la clausura de  $S$ . Dado que  $C'$  es acotada y no decreciente en cada coordenada, podemos extender  $C'$  a una función  $C''$  con dominio  $\overline{S_1} \times \overline{S_2}$  tomando límites. Claramente  $C''$  es basada y  $C''(1, v) = v$ ,  $C''(u, 1) = u$ , para todo  $(u, v) \in \overline{S_1} \times \overline{S_2}$ . Si tomamos un rectángulo  $\overline{B}$  en  $\overline{S_1} \times \overline{S_2}$ , cada vértice de  $\overline{B}$  es límite de una sucesión de puntos en  $S_1 \times S_2$ , y sea  $B_n$  una sucesión de rectángulos en  $S_1 \times S_2$ , tal que el límite de los vértices de  $B_n$  sean los vértices de  $\overline{B}$ . Por definición de  $C''$ , tenemos  $V_{C''}(\overline{B}) = \lim_{n \rightarrow +\infty} V_{C'}(B_n) \geq 0$ , y por tanto  $C''$  es 2-creciente. Concluimos que  $C''$  es una subcópula.

Por último, debemos extender  $C''$  al resto de  $\mathbf{I}^2$ . Sea  $(a, b) \in \mathbf{I}^2$ . Sean  $a_1 = \max\{u \in \overline{S_1} \mid u \leq a\}$  y  $a_2 = \min\{u \in \overline{S_1} \mid u \geq a\}$ , y de manera análoga definimos  $b_1$  y  $b_2$ . Por ejemplo, si  $a \in \overline{S_1}$ , entonces  $a_1 = a = a_2$ . Definamos

$$\lambda_1 = \begin{cases} \frac{a-a_1}{a_2-a_1} & \text{si } a_2 \neq a_1 \\ 1 & \text{si } a_1 = a_2 \end{cases}$$

$$\mu_1 = \begin{cases} \frac{b-b_1}{b_2-b_1} & \text{si } b_2 \neq b_1 \\ 1 & \text{si } b_1 = b_2 \end{cases}$$

y con esto podemos definir una cópula  $C: \mathbf{I}^2 \rightarrow \mathbf{I}$  como

$$C(a, b) = (1 - \lambda_1)(1 - \mu_1)C''(a_1, b_1) + (1 - \lambda_1)\mu_1C''(a_1, b_2) + \lambda_1(1 - \mu_1)C''(a_2, b_1) + \lambda_1\mu_1C''(a_2, b_2).$$

Claramente  $C|_{\overline{S_1} \times \overline{S_2}} = C''$ . Por lo tanto, para ver que  $C$  es una cópula, solo basta verificar que es 2-creciente. Sea  $(c, d) \in \mathbf{I}^2$  tal que  $(a, b), (c, d)$ , con  $a \leq c, b \leq d$ , son los vértices opuestos que definen un rectángulo  $B \subset \mathbf{I}^2$ . Sean  $c_1, c_2, d_1, d_2, \lambda_2, \mu_2$  los valores relacionados a  $C(c, d)$ . Para calcular  $V_C(B)$ , resulta importante saber si hay puntos de  $\overline{S_1}$  estrictamente entre  $a$  y  $c$ , y si hay puntos de  $\overline{S_2}$  estrictamente entre  $b$  y  $d$ . En el caso más sencillo eso no sucede, por lo tanto  $a_i = c_i, b_i = d_i$  con  $i = 1, 2$ , y se tiene

$$\begin{aligned} V_C(B) &= C(c, d) - C(c, b) - C(a, d) + C(a, b) \\ &= C''(a_1, b_1)[(1 - \lambda_2)(1 - \mu_2) - (1 - \lambda_2)(1 - \mu_1) - \\ &\quad - (1 - \lambda_1)(1 - \mu_2) + (1 - \lambda_2)(1 - \mu_1)] \\ &\quad + C''(a_1, b_2)[(1 - \lambda_2)\mu_2 - (1 - \lambda_2)\mu_1 - (1 - \lambda_1)\mu_2 + (1 - \lambda_1)\mu_1] \\ &\quad + C''(a_2, b_1)[\lambda_2(1 - \mu_2) - \lambda_2(1 - \mu_1) - \lambda_1(1 - \mu_2) + \lambda_1(1 - \mu_1)] \\ &\quad + C''(a_2, b_2)[\lambda_2\mu_2 - \lambda_2\mu_1 - \lambda_1\mu_2 + \lambda_1\mu_1] \\ &= (\lambda_2 - \lambda_1)(\mu_2 - \mu_1)V_{C''}(B) \\ &\geq 0 \end{aligned}$$

Supongamos ahora que  $a < a_2 \leq c_1 < c, b < b_2 \leq d_1 < d$ . Aplicando la

definición de  $V_C(B)$  podemos verificar que

$$\begin{aligned}
V_C(B) = & (1 - \lambda_1)\mu_2 V_{C''}([a_1, a_2] \times [d_1, d_2]) + \mu_2 V_{C''}([a_2, c_1] \times [d_1, d_2]) \\
& + \lambda_2 \mu_2 V_{C''}([c_1, c_2] \times [d_1, d_2]) + (1 - \lambda_1) V_{C''}([a_1, a_2] \times [b_2, d_1]) \\
& + V_{C''}([a_2, c_1] \times [b_2, d_1]) + \lambda_2 V_{C''}([c_1, c_2] \times [b_2, d_1]) \\
& + (1 - \lambda_1)(1 - \mu_1) V_{C''}([a_1, a_2] \times [b_1, b_2]) \\
& + (1 - \mu_1) V_{C''}([a_2, c_1] \times [b_1, b_2]) \\
& + \lambda_2(1 - \mu_1) V_{C''}([c_1, c_2] \times [b_1, b_2])
\end{aligned}$$

y como todos los términos son positivos, se tiene que en este caso,  $V_C(B) \geq 0$ . El resto de la discusión se resuelve de la misma manera.  $\square$

La demostración del teorema de Sklar [1.2.4] es inmediata a partir de estos dos lemas.

*Demostración.* Demostración del teorema de Sklar

Dada  $H$ , a partir del lema 1.2.5, obtenemos una subcópula  $C'$  definida unívocamente en  $\text{Ran}F \times \text{Ran}G$ . Si  $F$  y  $G$  son continuas, entonces  $\text{Ran}F \times \text{Ran}G = \mathbf{I}^2$ , de lo contrario, podemos extender  $C'$  a una cópula  $C$  utilizando el lema 1.2.6.

Para el recíproco, sea  $H(x, y) = C(F(x), G(y))$ .  $H$  es basada por serlo  $C$ , y es 2-creciente por ser  $F$  y  $G$  no decrecientes, y  $C$  ser 2-creciente. Además  $H(+\infty, +\infty) = C(1, 1) = 1$ .  $\square$

Además, la fórmula del teorema de Sklar -que expresa la función de distribución conjunta en función de una cópula y las dos marginales- puede ser invertida. Luego de invertida, la fórmula muestra como escribir la cópula en función de la distribución conjunta (y sus marginales). Para lograr la inversión de la fórmula del teorema de Sklar, necesitamos poder “invertir” funciones de distribución no biyectivas.

**Definición 1.2.7.** Sea  $F$  una función de distribución. Una **cuasi-inversa** de  $F$  es una función  $F^{(-1)}$  con dominio  $\mathbf{I}$ , tal que

- Si  $t \in \text{Ran}F$ , entonces  $F^{(-1)}(t)$  es un número  $x \in \mathbb{R}$  tal que  $F(x) = t$ , o sea,  $F(F^{(-1)}(t)) = t$ .
- Si  $t \notin \text{Ran}F$ , entonces  $F^{(-1)}(t) = \inf\{x \mid F(x) \geq t\} = \sup\{x \mid F(x) \leq t\}$

*Observación 1.2.8.* Si  $F$  es estrictamente creciente, entonces la cuasi-inversa  $F^{(-1)}$  es única. Si además  $F: \overline{\mathbb{R}} \rightarrow \mathbf{I}$  es biyectiva, entonces la cuasi-inversa  $F^{(-1)}$  coincide con  $F^{-1}$ , la función inversa de  $F$ .

Con esta notación, podemos enunciar el siguiente lema:

**Lema 1.2.9.** *Sean  $H$ ,  $F$ ,  $G$  y  $C'$  como en el lema 1.2.5, y sean  $F^{(-1)}$ ,  $G^{(-1)}$  cuasi-inversas de  $F$  y  $G$  respectivamente. Entonces, para todo  $(u, v) \in \text{Dom}C'$ , se tiene*

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

En el caso que  $F$  y  $G$  sean continuas, el dominio de  $C'$  es  $\mathbf{I}^2$ . Por lo tanto, el lema anterior puede pensarse como una herramienta para construir cópulas: si tomamos una función de distribución conjunta  $H$ , podemos hallarle las marginales  $F$  y  $G$ ; si estas son contínuas, podemos hallar una función  $C$  asociada a partir de la fórmula del lema anterior, y dicha  $CF$  necesariamente será una cópula.

### 1.3. Vectores aleatorios y cópulas.

Sea  $(\Omega, \mathfrak{A}, \mathbb{P})$  un espacio de probabilidad, y sean  $X, Y: \Omega \rightarrow \mathbb{R}$  variables aleatorias. Llamemos  $F_X$  y  $F_Y$  a las funciones de distribución de  $X$  e  $Y$  respectivamente; dichas funciones serán funciones de distribución (según la definición de la primera sección de este capítulo), y además serán continuas por derecha. Cuando hablemos del vector  $(X, Y): \Omega \rightarrow \mathbb{R}^2$  nos referiremos a un vector aleatorio con distribución conjunta  $F_{XY}$ ; dicha función será una distribución conjunta (en el sentido de la definición de la primera sección de este capítulo), y tendrá marginales  $F_X$  y  $F_Y$ .

Con esta notación podemos enunciar el teorema de Sklar en términos probabilísticos:

**Teorema 1.3.1.** *Teorema de Sklar*

*Sean  $X, Y$  variables aleatorias con funciones de distribución  $F_X$  y  $F_Y$  respectivamente, y función de distribución conjunta  $F_{XY}$ . Entonces existe una cópula  $C_{XY}$  tal que*

$$F_{XY}(x, y) = C_{XY}(F_X(x), F_Y(y))$$

*Si  $F_X$  y  $F_Y$  son continuas entonces  $C_{XY}$  es única. De lo contrario,  $C_{XY}$  está unívocamente determinada en  $\text{Ran}F_X \times \text{Ran}F_Y$*

**Definición 1.3.2.** Llamaremos **cópula de  $X$  e  $Y$**  a la cópula  $C_{XY}$  del teorema de Sklar.



*Observación 1.3.3.* Toda cópula  $C$  puede ser pensada -luego de extender su dominio a  $\overline{\mathbb{R}}^2$  convenientemente- como la función de distribución de un vector aleatorio  $(U, V)$  donde  $U$  y  $V$  son variables aleatorias uniformes  $(0, 1)$ . La extensión es  $F_{UV}(x, y) = C(\hat{x}, \hat{y})$  donde

$$\hat{x} = \begin{cases} 0 & \text{si } x \leq 0 \\ x & \text{si } x \in (0, 1) \\ 1 & \text{si } x \geq 1 \end{cases}$$

y de manera análoga se define  $\hat{y}$ .

El primer resultado interesante acerca de cópulas de variables aleatorias surge al preguntarse que pasa en el caso que  $X$  e  $Y$  sean independientes.

**Teorema 1.3.4.** Sean  $X, Y$  variables aleatorias con funciones de distribución  $F_X$  y  $F_Y$  continuas. Entonces  $X$  e  $Y$  son independientes si y solo si  $C_{XY}(u, v) = uv$ .

*Demostración.* Supongamos que  $X$  e  $Y$  son independientes. Entonces

$$F_{XY}(x, y) = F_X(x)F_Y(y).$$

Entonces, por el lema 1.2.9 se tiene que para todo  $(u, v) \in \mathbf{I}^2$ ,  $C_{XY}(u, v) = F_{XY}(F_X^{(-1)}(u), F_Y^{(-1)}(v)) = uv$ .

Análogamente, si  $C_{XY}(u, v) = uv$ , entonces por el teorema de Sklar,  $F_{XY}(x, y) = C_{XY}(F_X(x), F_Y(y)) = F_X(x)F_Y(y)$ .  $\square$

**Notación.** Llamaremos  $\Pi(u, v) = uv$  a la cópula que caracteriza la independencia de variables aleatorias.

*Observación 1.3.5.* La hipótesis de continuidad de  $F_X$  y  $F_Y$  se pide únicamente para la unicidad de la cópula. Se podría enunciar el teorema anterior en base a la subcópula  $C'_{XY}$  del lema 1.2.5. De esa manera, sin pedir continuidad de las marginales, obtendríamos que  $X$  e  $Y$  son independientes si y solo si  $C'_{XY}(u, v) = uv$ .

Otro ejemplo interesante, en este caso de dependencia extrema, está relacionado con las cotas de Fréchet-Hoeffding.

**Lema 1.3.6.** Sea  $X$  una variable aleatoria con función de distribución  $F_X$  continua, y  $(X, X)$  el vector aleatorio que surge de poner en ambas variables el mismo valor  $X(\omega)$ , entonces  $C_{XX}(u, v) = M(u, v)$ , donde  $M$  es la cota superior de Fréchet-Hoeffding.

*Demostración.* La idea es aplicar el lema 1.2.9, igual que en el teorema anterior. Para eso debemos calcular  $F_{XX}$ :

$$F_{XX}(x, y) = \mathbb{P}\{X \leq x, X \leq y\} = \mathbb{P}\{X \leq \min\{x, y\}\} = F_X(\min\{x, y\})$$

Por lo tanto, obtenemos

$$\begin{aligned} C_{XX}(u, v) &= F_{XX}(F_X^{(-1)}(u), F_X^{(-1)}(v)) \\ &= F_X(\min\{F_X^{(-1)}(u), F_X^{(-1)}(v)\}) \\ &= \min\{u, v\} = M(u, v). \end{aligned}$$

□

Enunciaremos a continuación un teorema que utilizaremos mucho en nuestro modelo. La idea es que, si tenemos un vector aleatorio  $(X, Y)$ , y aplicamos transformaciones monótonas estrictamente crecientes (llamémoslas  $\alpha$  y  $\beta$ ) a cada coordenada para obtener el nuevo vector aleatorio  $(\alpha(X), \beta(Y))$ , entonces las cópulas del teorema de Sklar  $C_{XY}$  y  $C_{\alpha(X)\beta(Y)}$  serán iguales. En nuestro caso, esto nos servirá porque no tendremos observaciones directas de la variable aleatoria que nos interesa, pero sí tendremos observaciones que podremos relacionar a nuestra variable aleatoria de interés via una función creciente (ver 4.1.4).

**Teorema 1.3.7.** *Sea  $(X, Y)$  un vector aleatorio, cuyas marginales  $X$  e  $Y$  son variables aleatorias con funciones de distribución  $F_X$  y  $F_Y$  continuas. Sean  $\alpha, \beta$  funciones estrictamente crecientes definidas en  $\text{Ran}X$  y  $\text{Ran}Y$  respectivamente, y consideremos el nuevo vector aleatorio  $(\alpha(X), \beta(Y))$ . Entonces  $C_{XY} = C_{\alpha(X)\beta(Y)}$ .*

*Demostración.* Se tiene

$$\begin{aligned} C_{\alpha(X), \beta(Y)}(F_{\alpha(X)}(x), F_{\beta(Y)}(y)) &= \mathbb{P}\{\alpha(X) \leq x, \beta(Y) \leq y\} \\ &= \mathbb{P}\{X \leq \alpha^{-1}(x), Y \leq \beta^{-1}(y)\} \\ &= C_{XY}(F_X(\alpha^{-1}(x)), F_Y(\beta^{-1}(y))) \\ &= C_{XY}(F_{\alpha(X)}(x), F_{\beta(Y)}(y)). \end{aligned}$$

Entonces  $C_{\alpha(X), \beta(Y)} = C_{XY}$  en  $\text{Ran}F_{\alpha(X)} \times \text{Ran}F_{\beta(Y)} = \mathbf{I}^2$  por ser  $F_X, F_Y$  continuas y  $\alpha, \beta$  estrictamente crecientes. □

El teorema 1.3.7 se puede extender a transformaciones estrictamente decrecientes. Dado que las demostraciones son análogas, enunciaremos el teorema sin demostración.

**Teorema 1.3.8.** *Teorema 1.3.7 extendido*

Sea  $(X, Y)$  un vector aleatorio, cuyas marginales  $X$  e  $Y$  son variables aleatorias con funciones de distribución  $F_X$  y  $F_Y$  continuas, y sean  $\alpha$  y  $\beta$  funciones definidas en  $\text{Ran}X$  y  $\text{Ran}Y$  respectivamente.

- Si  $\alpha$  es estrictamente creciente, y  $\beta$  es estrictamente decreciente, entonces  $C_{\alpha(X)\beta(Y)}(u, v) = u - C_{XY}(u, 1 - v)$ . El resultado es análogo intercambiando los roles de  $\alpha$  y  $\beta$ .
- Si  $\alpha$  y  $\beta$  son estrictamente decrecientes, entonces  $C_{\alpha(X)\beta(Y)}(u, v) = u + v - 1 + C_{XY}(1 - u, 1 - v)$ .

*Observación 1.3.9.* Supongamos que estamos en las hipótesis del lema 1.3.6, donde  $X$  es una variable aleatoria con  $F_X$  continua, y  $(X, X)$  el vector aleatorio con ambas coordenadas iguales a  $X(\omega)$ .

- Sea  $\beta$  estrictamente creciente, y apliquémosla a la segunda coordenada. Via el teorema 1.3.7, tenemos que el vector aleatorio de la forma  $(X, \beta(X))$  tiene asociada la cópula  $M(u, v)$ .
- Si  $\beta(x) = -x$  (o cualquier función decreciente), obtenemos también via el teorema 1.3.7, que el vector aleatorio de la forma  $(X, -X)$  tiene asociada la cópula  $C_{X-X}(u, v) = u - M(u, v) = \min\{u - u, u - 1 + v\} = W(u, v)$ .

## 1.4. Medidas clásicas de dependencia

Ya dijimos que las cópulas pueden ser pensadas como una función que retiene la estructura de dependencia. Esto induce a pensar que podemos relacionar medidas clásicas de dependencia invariantes por escala (como la *rho de Spearman* o la *tau de Kendall*) con las cópulas. Evidentemente, las medidas de dependencia que no sean invariantes por escala (por ejemplo, la correlación lineal) no puede extraerse de una cópula, ya que depende de las marginales.

El concepto que miden, tanto la rho de Spearman como la tau de Kendall, es la concordancia. Sea  $(X, Y)$  un vector aleatorio con marginales continuas, y  $(X_1, Y_1), (X_2, Y_2)$  dos observaciones. Decimos que las observaciones son

**concordantes** si  $(X_2 - X_1)(Y_2 - Y_1) > 0$ , y **discordantes** si  $(X_2 - X_1)(Y_2 - Y_1) < 0$ . Claramente, el concepto de concordancia no depende de la escala.

#### 1.4.1. Tau de Kendall

Supongamos que tenemos una muestra aleatoria simple  $\mathbb{M}_n = \{(X_i, Y_i), i = 1, \dots, n\}$ , con  $(X_i, Y_i) \sim (X, Y) \in \mathbb{R}^2$ , y las marginales de la distribución de  $(X, Y)$  son continuas. Tendremos entonces que podemos formar  $\binom{n}{2}$  pares distintos con elementos de  $\mathbb{M}_n$ , algunos de ellos serán concordantes (denotaremos por  $c$  a la cantidad de pares concordantes), los demás discordantes ( $d$  pares discordantes). Definimos entonces

$$t = \frac{c - d}{c + d}$$

la proporción de pares concordantes menos pares discordantes. Si  $(X_i, Y_i)$  y  $(X_j, Y_j)$  son dos observaciones independientes de  $(X, Y)$ , podemos pensar a  $t$  como la versión muestral de

$$\tau_{XY} = \mathbb{P}\{(X_i - X_j)(Y_i - Y_j) > 0\} - \mathbb{P}\{(X_i - X_j)(Y_i - Y_j) < 0\}$$

o sea, la probabilidad de concordancia menos la probabilidad de discordancia. Llamamos a  $\tau_{XY}$  la **tau de Kendall**.

Veamos ahora que  $\tau_{XY}$  solo depende de  $C_{XY}$  demostrando esta proposición más general, que estudia la concordancia entre dos muestras provenientes de distintas funciones de distribución conjuntas.

**Proposición 1.4.1.** *Sea  $\mathbb{M}_n^1 = \{(X_i^1, Y_i^1)\}$  una observación proveniente de la distribución  $H_1$ , y  $\mathbb{M}_n^2 = \{(X_j^2, Y_j^2)\}$  una observación proveniente de la distribución  $H_2$ , ambas leyes con las mismas funciones de distribución marginales (continuas), y supongamos que ambas observaciones se toman independientes. Sean  $C_1$  la cópula de  $H_1$  y  $C_2$  la cópula de  $H_2$ . Entonces la probabilidad de concordancia menos la probabilidad de discordancia entre  $(X_i^1, Y_i^1)$  y  $(X_j^2, Y_j^2)$  vale*

$$4 \iint_{\mathcal{I}^2} C_2(u, v) dC_1(u, v) - 1$$

*Demostración.* Como las marginales de  $H_1$  y  $H_2$  son continuas, entonces la probabilidad de empate  $(X_i = X_j, Y_i = Y_j)$  son nulas. Por lo tanto

$$\mathbb{P}\{(X_i - X_j)(Y_i - Y_j) > 0\} = 1 - \mathbb{P}\{(X_i - X_j)(Y_i - Y_j) < 0\}$$

Además, tenemos

$$\mathbb{P}\{(X_i - X_j)(Y_i - Y_j) > 0\} = \mathbb{P}\{(X_i - X_j) > 0, (Y_i - Y_j) > 0\} + \mathbb{P}\{(X_i - X_j) < 0, (Y_i - Y_j) < 0\}$$

y ambas probabilidades pueden ser escritas en función de  $C$ :

$$\begin{aligned} \mathbb{P}\{(X_i - X_j) > 0, (Y_i - Y_j) > 0\} &= \mathbb{P}\{X_j \leq X_i, Y_j \leq Y_i\} \\ &= \iint_{\mathbb{R}^2} \mathbb{P}\{X_j \leq x, Y_j \leq y\} dC_1(F_X(x), F_Y(y)) \\ &= \iint_{\mathbb{R}^2} C_2(F_X(x), F_Y(y)) dC_1(F_X(x), F_Y(y)) \\ &= \iint_{\mathbf{I}^2} C_2(u, v) dC_1(u, v) \end{aligned}$$

y de manera similar

$$\begin{aligned} \mathbb{P}\{(X_i - X_j) < 0, (Y_i - Y_j) < 0\} &= \iint_{\mathbf{I}^2} (1 - u - v + C_2(u, v)) dC_1(u, v) \\ &= 1 - \frac{1}{2} - \frac{1}{2} + \iint_{\mathbf{I}^2} C_2(u, v) dC_1(u, v) \\ &= \iint_{\mathbf{I}^2} C_2(u, v) dC_1(u, v) \end{aligned}$$

de lo cual concluimos  $\mathbb{P}\{(X_i - X_j)(Y_i - Y_j) > 0\} = 2 \iint_{\mathbf{I}^2} C_2(u, v) dC_1(u, v)$ .  
□

*Observación 1.4.2.* En el caso particular de la tau de Kendall,  $C_1 = C_2$ , y tenemos  $\tau_C = \iint_{\mathbf{I}^2} C(u, v) dC(u, v)$ .

Otra expresión para  $\tau_C$  es la siguiente ([1] pag 164, teorema 5.1.5):

**Teorema 1.4.3.**  $\iint_{\mathbf{I}^2} C(u, v) dC(u, v) = \frac{1}{2} - \iint_{\mathbf{I}^2} \frac{\partial C}{\partial u}(u, v) \frac{\partial C}{\partial v} C(u, v) dudv$ .

## 1.4.2. Rho de Spearman

Consideramos ahora otra idea para medir la concordancia. Consideremos tres observaciones independientes  $(X_i, Y_i)$ ,  $i = 1, 2, 3$ , provenientes del vector aleatorio  $(X, Y) \in \mathbb{R}^2$  con función de distribución  $F_{XY}$  y marginales  $F_X, F_Y$  continuas. Definimos la **Rho de Spearman** a una constante por la probabilidad de concordancia menos la probabilidad de discordancia entre  $(X_1, Y_1)$  y  $(X_2, Y_3)$

$$\rho_{XY} = 3(\mathbb{P}\{(X_1 - X_2)(Y_1 - Y_3) > 0\} - \mathbb{P}\{(X_1 - X_2)(Y_1 - Y_3) < 0\})$$

Dicho de otra manera, tomamos una observación bajo la ley  $F_{XY}(x, y)$ , un vector bajo la ley  $F_X(x)F_Y(y)$ , y tomamos la probabilidad de concordancia menos la probabilidad de discordancia (y multiplicamos por 3). También puede usarse el vector  $(X_3, Y_2)$ .

Si la cópula de  $F_{XY}$  es  $C$ , y al no depender  $\rho_{XY}$  de la escala, podemos pensar que la primera observación proviene de la cópula  $C$  y la segunda observación de la cópula  $\Pi(u, v) = uv$ . Formalizamos esto en el siguiente resultado:

**Proposición 1.4.4.** *Sea  $(X, Y) \in \mathbb{R}^2$  un vector aleatorio con marginales continuas y cópula  $C$ . Entonces se tiene*

$$\begin{aligned}\rho_{XY} = \rho_C &= 12 \iint_{\mathbf{I}^2} uv \, dC(u, v) - 3 \\ &= 12 \iint_{\mathbf{I}^2} C(u, v) \, dudv - 3\end{aligned}$$

*Demostración.* Aplicando la proposición 1.4.1 a nuestro caso ( $C_1 = C$ ,  $C_2 = \Pi$ ) obtenemos

$$\rho_{XY} = 3 \left( 4 \iint_{\mathbf{I}^2} uv \, dC(u, v) \right).$$

Además, del enunciado de la proposición 1.4.1 vemos que los roles de  $H_1$  y  $H_2$  son intercambiables, por lo tanto si  $C_1 = \Pi$ ,  $C_2 = C$  obtenemos

$$\rho_{XY} = 3 \left( 4 \iint_{\mathbf{I}^2} C(u, v) \, dudv \right).$$

□

Supongamos que tenemos un vector aleatorio  $(X, Y)$  con marginales  $F_X$  y  $F_Y$ , y cópula  $C$ . Como  $\rho_C$  no depende de la escala, consideremos  $(F_X(X), F_Y(Y))$ , un vector aleatorio con marginales uniformes  $U, V$ , y función de distribución (cópula)  $C$ . Entonces

$$\begin{aligned}\rho_{F_X(X)F_Y(Y)} = \rho_C &= 12 \iint_{\mathbf{I}^2} uv \, dC(u, v) - 3 \\ &= 12\mathbb{E}(UV) - 3 = \frac{\mathbb{E}(UV) - 1/4}{1/12} \\ &= \frac{\mathbb{E}(UV) - \mathbb{E}(U)\mathbb{E}(V)}{\sqrt{\text{Var}(U)\text{Var}(V)}}\end{aligned}$$

donde usamos que la esperanza de una variable aleatoria uniforme  $U(0, 1)$  es  $1/2$ , y la varianza es  $1/12$ . Podemos concluir que  $\rho_C$  es la correlación lineal entre  $F_X(X)$  y  $F_Y(Y)$ .

Otra interpretación surge de notar que  $\iint_{\mathbf{I}^2} uv \, dudv = 1/4$ . Tenemos que

$$\rho_C = 12 \iint_{\mathbf{I}^2} C(u, v) \, dudv - 3 = 12 \left( \iint_{\mathbf{I}^2} [C(u, v) - uv] \, dudv \right)$$

y podemos pensar que  $\rho_C$  es proporcional a la diferencia (promediada) entre  $C$  y  $\Pi$ .

## 1.5. Dependencias de colas

En el contexto de dependencia, nos interesa saber si los grandes o pequeños valores de las coordenadas del vector aleatorio vienen acompañados. Más aún, en nuestro contexto de dependencia en el default bancario, justamente nos interesa la probabilidad de ocurrencia de valores bajos del activo de un banco  $X$  dados valores bajos del activo del banco  $Y$ . Las dependencias de colas son un concepto límite de dicha idea:

**Definición 1.5.1.** Sea  $(X, Y)$  un vector aleatorio, cuyas marginales  $X$  e  $Y$  son variables aleatorias con funciones de distribución  $F_X$  y  $F_Y$  y cópula  $C_{XY}$ . Definimos la **dependencia de cola superior**  $\lambda_U$  como

$$\lambda_U = \lim_{t \rightarrow 1^-} \mathbb{P} \left\{ Y \geq F_Y^{(-1)}(t) \mid X \geq F_X^{(-1)}(t) \right\}$$

De manera similar, definimos la **dependencia de cola inferior** al límite

$$\lambda_L = \lim_{t \rightarrow 0^+} \mathbb{P} \left\{ Y \leq F_Y^{(-1)}(t) \mid X \leq F_X^{(-1)}(t) \right\}$$

Informalmente, podemos definir la dependencia de cola superior como el límite de la probabilidad que  $Y$  sea superior al percentil  $100t$  sabiendo que  $X$  es superior al percentil  $100t$ , con  $t$  tendiendo a 1 (por la izquierda). Análogo para la dependencia de cola inferior.

Se puede ver que la existencia y el valor de las dependencias de colas solo dependen de la cópula del vector aleatorio  $(X, Y)$ .

**Proposición 1.5.2.** *Supongamos las hipótesis de la definición 1.5.1, y que la dependencia de colas  $\lambda_U$  y  $\lambda_L$  existe. Entonces vale*

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

*Demostración.* Haremos las cuentas para  $\lambda_U$ , ya que para  $\lambda_L$  es análogo.

$$\begin{aligned}
\lambda_U &= \lim_{t \rightarrow 1^-} \mathbb{P} \left\{ Y \geq F_Y^{(-1)}(t) \mid X \geq F_X^{(-1)}(t) \right\} \\
&= \lim_{t \rightarrow 1^-} \mathbb{P} \{ F_Y(Y) \geq t \mid F_X(X) \geq t \} \\
&= \lim_{t \rightarrow 1^-} \frac{1 - t - t + C_{XY}(t, t)}{1 - t} \\
&= 2 - \lim_{t \rightarrow 1^-} \frac{1 - C_{XY}(t, t)}{1 - t}.
\end{aligned}$$

□

Acotando el valor de la cópula  $C_{XY}(t, t)$  por las cotas de Fréchet-Hoeffding, podemos acotar los posibles valores de  $\lambda_U$  y  $\lambda_L$ .

**Proposición 1.5.3.** *En las hipótesis de la definición 1.5.1, si las dependencias de colas  $\lambda_U$  y  $\lambda_L$  existen, entonces  $0 \leq \lambda_U, \lambda_L \leq 1$ .*

*Demostración.* Haremos las cuentas solo para la dependencia de colas superior, ya que para la dependencia de colas inferior es análogo. Veamos primero que  $\lambda_U \leq 1$ :

$$\begin{aligned}
\lambda_U &= 2 - \lim_{t \rightarrow 1^-} \frac{1 - C_{XY}(t, t)}{1 - t} \\
&\leq 2 - \frac{1 - M(t, t)}{1 - t} \\
&\leq 2 - \frac{1 - t}{1 - t} = 1.
\end{aligned}$$

Veamos que  $\lambda_U \geq 0$

$$\begin{aligned}
\lambda_U &= 2 - \lim_{t \rightarrow 1^-} \frac{1 - C_{XY}(t, t)}{1 - t} \\
&\geq 2 - \frac{1 - W(t, t)}{1 - t} \\
&\geq 2 - \frac{2 - 2t}{1 - t} = 0
\end{aligned}$$

□



Un resultado sencillo referente a las dependencias de colas puede verse a través de la densidad de la cópula (en los casos que exista).

**Proposición 1.5.4.** *Supongamos las hipótesis de la definición 1.5.1, y además que existe la densidad de la cópula,  $c_{XY}(u, v) = \frac{\partial^2 C_{XY}(u, v)}{\partial v \partial u}$ . Si existe  $K \in \mathbb{R}$  tal que  $c_{XY}(u, v) \leq K$  entonces  $\lambda_U = \lambda_L = 0$*

*Demostración.* Haremos las cuentas solo para la dependencia de colas superior, ya que para la dependencia de colas inferior es análogo.

$$\begin{aligned} \lambda_U &= 2 - \lim_{t \rightarrow 1^-} \frac{1 - C_{XY}(t, t)}{1 - t} \\ &= \lim_{t \rightarrow 1^-} \frac{1 - 2t + C_{XY}(t, t)}{1 - t} \\ &\leq \lim_{t \rightarrow 1^-} \frac{K(1 - t)^2}{1 - t} \\ &= \lim_{t \rightarrow 1^-} K(1 - t) = 0 \end{aligned}$$

La desigualdad se da por el hecho que  $1 - 2t + C_{XY}(t, t) = V_C([t, 1]^2) = \iint_{[t, 1]^2} c_{XY}(x, y) dy dx \leq K(1 - t)^2$ .  $\square$

*Observación 1.5.5.* En la proposición anterior, no es necesario que toda la densidad esté acotada para que esté acotada la dependencia de cola superior. Lo que realmente importa es que la densidad esté acotada en  $[t, 1]^2$  para algún  $t$ .

Cuando trabajamos con una muestra aleatoria simple (por ejemplo en nuestro caso), es sumamente complicado evaluar la dependencia de cópulas, ya que se trata de un límite en valores extremos de la cópula: no tendremos datos en la muestra mayores a un cuantil suficientemente cercano a 100. Más aún, se pueden simular datos provenientes de una cópula con dependencias de colas nulas pero que presenten indicios de dependencia de colas alta. Es por eso que solo estudiaremos la dependencia de colas cuando hablemos de estimación paramétrica de colas: un buen ajuste en una familia de cópulas que presente dependencias de colas será el mejor indicio de dependencias de colas al que podemos aspirar.

## Capítulo 2

# Estimación no paramétrica de cópulas

La estimación no paramétrica de cópulas o densidades de cópulas (o la estimación no paramétrica de densidades de distribuciones en general) es un problema de gran interés en la estadística. No importa qué muestra aleatoria simple tengamos, casi no necesitamos hacer suposiciones (como mucho marginales continuas, o suponer que exista densidad de la función de distribución conjunta) para poder obtener una estimación de la cópula o de su densidad.

En dimensión 1, los ejemplos más conocidos de estimación no paramétrica de densidades son el histograma y la estimación por kernels gaussianos (ver [8]). Sin embargo, ambas metodologías llevadas al problema de estimación de densidades de cópulas parecen tener grandes desventajas (las cuales hemos comprobado a través de simulaciones utilizando el software **R**, usando tamaños de muestra comparables a los tamaños de muestra que manejaremos más adelante, esto es, alrededor de 500, ver 4.1.4).

En el caso de los histogramas, tomar grillas muy finas lleva a gráficas muy puntiagudas, y tomar grillas gruesas lleva a dibujos sin muchos cambios, que brindan poca información. Peor aún, no siempre se observan los máximos en los lugares del cuadrado  $\mathbf{I}^2$  donde deberían estar, lo cual puede llevar a problemas de interpretación, y dificultad de conjeturar posibles familias de cópulas para estimaciones paramétricas.

En el caso de los kernels gaussianos, tenemos primero problemas conceptuales. El primero y más evidente, la idea de kernels gaussianos no se adapta bien a soportes compactos. El segundo tiene que ver con que la estimación por kernels gaussianos pretende ser independiente de la elección de

los ejes de coordenadas: esto es, no hay información de utilidad a priori en los ejes, y utilizamos kernels que se ajusten a la elección de ejes de nuestra preferencia. En nuestro caso es totalmente lo contrario: los ejes coordenados representarán las variables aleatorias  $X$  e  $Y$  (eventualmente representarán a nuestros datos bancarios), de las cuales siempre queremos tener referencia. Y de vuelta tenemos el mismo problema que con el histograma: no es fácil identificar donde se encuentran los máximos y mínimos de la densidad verdadera.

Esto nos inspira a buscar maneras alternativas de estimar las cópulas y sus densidades de manera no paramétrica. En la primera sección estudiaremos la cópula empírica (un análogo a la distribución empírica) y sus utilidades. En la segunda sección proponemos una estimación no paramétrica de densidades de cópulas basado en el método conocido como “random forests” propuesto por Breiman para el problema de clasificación supervisada (ver [2] para resultados de consistencia de estos estimadores). En la tercera sección estudiaremos algunas propiedades de este estimador, por ejemplo la consistencia del estimador, y en la cuarta sección mostramos simulaciones en el software **R** de la densidad usando random forests; simulamos datos provenientes de una cópula Gumbel (ver 3.1) y veremos los resultados obtenidos para saber qué podemos esperar del estimador.

## 2.1. Cópula empírica

Esta sección está basada en ([1] sección 5.6).

Partiendo de una muestra aleatoria simple  $\mathbb{X}_n = (X_1, Y_1), \dots, (X_n, Y_n)$ ,  $(X_i, Y_i) \in \mathbb{R}^2$ , de un vector aleatorio  $(X, Y)$ , la primera idea para la construcción no paramétrica de la cópula correspondiente al vector aleatorio  $X$  es la misma que usamos para construir las funciones de distribución empíricas.

**Definición 2.1.1.** Sea  $\mathbb{X}_n$  una muestra aleatoria simple de tamaño  $n$  del vector aleatorio  $(X, Y)$ .

- Llamamos **cópula empírica**  $C_n$  a la función definida como

$$C_n \left( \frac{i}{n}, \frac{j}{n} \right) = \frac{\#\{(X_k, Y_k) \in \mathbb{X}_n \mid X_k \leq X_{(i)}, Y_k \leq Y_{(j)}\}}{n}$$

donde  $X_{(i)}$  y  $Y_{(j)}$  denotan los estadísticos de orden de  $\mathbb{X}_n$ .

- Definimos la **frecuencia de la cópula empírica**  $c_n$  como

$$c_n \left( \frac{i}{n}, \frac{j}{n} \right) = \begin{cases} \frac{1}{n} & \text{si } (X_{(i)}, Y_{(j)}) \in \mathbb{X}_n \\ 0 & \text{en otro caso} \end{cases}$$

donde  $X_{(i)}$  y  $Y_{(j)}$  denotan los estadísticos de orden de  $\mathbb{X}_n$ .

*Observación 2.1.2.* Se cumplen las siguientes igualdades:

- $C_n \left( \frac{i}{n}, \frac{j}{n} \right) = \sum_{a=1}^i \sum_{b=1}^j c_n \left( \frac{a}{n}, \frac{b}{n} \right)$
- $c_n \left( \frac{i}{n}, \frac{j}{n} \right) = C_n \left( \frac{i}{n}, \frac{j}{n} \right) - C_n \left( \frac{i-1}{n}, \frac{j}{n} \right) - C_n \left( \frac{i}{n}, \frac{j-1}{n} \right) + C_n \left( \frac{i-1}{n}, \frac{j-1}{n} \right)$

Una utilidad de la cópula empírica y de la frecuencia empírica es para hallar la versión muestral de la tau de Kendall y de la rho de Spearman ([1], página 220 teorema 5.6.2).

*Proposición 2.1.3.* Sean  $t$  y  $r$  las versiones muestrales de la tau de Kendall y la rho de Spearman respectivamente, y sea  $C$  una cópula. Dada una muestra  $\mathbb{M}_n \subset \mathbf{I}^2$  proveniente de la cópula  $C$ , denotaremos por  $C_n$  a la cópula empírica y por  $c_n$  a la frecuencia empírica. Entonces

$$t = \frac{2n}{n+1} \sum_{i,j=2}^n \sum_{a=1}^{i-1} \sum_{b=1}^{j-1} \left( c_n \left( \frac{i}{n}, \frac{j}{n} \right) c_n \left( \frac{a}{n}, \frac{b}{n} \right) - c_n \left( \frac{i}{n}, \frac{b}{n} \right) c_n \left( \frac{a}{n}, \frac{j}{n} \right) \right),$$

$$r = \frac{12}{n^2-1} \sum_{i=1}^n \sum_{j=1}^n \left( C_n \left( \frac{i}{n}, \frac{j}{n} \right) - \frac{ij}{n^2} \right).$$

## 2.2. Estimación de densidad por random forests

En esta sección proponemos un método no paramétrico para la estimación de cópulas bivariadas consideradas como densidades en  $\mathbf{I}^2$ . Dicho método se inspira en el método de bosques aleatorios (random forests) utilizado para el problema de clasificación supervisada [2]. Nos referiremos al estimador como **estimador de densidad por random forests**.

A continuación describiremos y construiremos el estimador de densidad por random forest. De ahora en más denotaremos por  $\mathbb{M}_n = (X_1, Y_1), \dots, (X_n, Y_n)$  con  $(X_i, Y_i) \in \mathbf{I}^2$  a una muestra aleatoria simple proveniente de una cópula, soportada en  $\mathbf{I}^2$ , con densidad  $p$ . Lo primero que haremos es generar una partición aleatoria por rectángulos de  $\mathbf{I}^2$  bajo el siguiente algoritmo:

1. La partición inicial de  $\mathbf{I}^2$  es  $\{\mathbf{I}^2\}$ . Lo fijamos como rectángulo inicial.
2. Se sortea al azar una de las dos coordenadas, ambas con probabilidad  $1/2$ .
3. Sorteamos un punto uniformemente en el lado del rectángulo fijado correspondiente a la coordenada elegida.
4. Partimos el rectángulo fijado con un segmento perpendicular a la coordenada seleccionada en el paso 2, a través del punto seleccionado en el paso 3. A partir de aquí tendremos un refinamiento de la partición: al partir uno de los rectángulos en dos, tenemos un rectángulo más que antes.
5. Sorteamos uniformemente uno de los rectángulos de la partición (total), lo fijamos, y volvemos al paso 2.

El algoritmo se repite  $k$  veces, donde  $k$  es un parámetro fijado por el usuario. Al final obtendremos una partición  $\mathfrak{D}_k$  de  $\mathbf{I}^2$ , donde dos rectángulos distintos de la partición son disjuntos, o bien su intersección es un lado común, o bien su intersección es un vértice común. Como la MAS proviene de una distribución con densidad, entonces la probabilidad de que una de las observaciones pertenezca al lado de uno de los rectángulos es cero.

**Notación.** Denotaremos por  $x$  a un punto en  $\mathbf{I}^2$ , y por  $\lambda$  a la medida de Lebesgue; al rectángulo de  $\mathfrak{D}_k$  (también llamado “celda”) que contiene a  $x$  lo denotaremos por  $A(x)$ ; a la cantidad de observaciones contenidas en  $A(x)$  la denotaremos por  $N(x)$  (claramente  $0 \leq N(x) \leq n$ ).

Por lo tanto podemos definir el **tree estimator**  $\hat{p}$  (que depende de  $\mathfrak{D}_k$ ) como

$$\hat{p}(x) = \frac{N(x)}{n\lambda(A(x))}$$

y, por último, definimos el **random forest estimator** como

$$p_k(x) = \mathbb{E}_{\mathfrak{D}_k}(\hat{p}(x))$$

donde  $E_{\mathfrak{D}_k}$  denota la esperanza al tomar todas las posibles particiones de  $k$  cortes.

El primer lema que veremos trata sobre el tamaño de las celdas de  $\mathfrak{D}_k$  cuando  $k$  tiende a infinito: más adelante, necesitaremos decir que, para  $k$  suficientemente grande, las celdas tendrán diámetro chico. Este teorema está demostrado en [2], pero también mostraremos aquí la demostración.

**Lema 2.2.1.** *Sea  $X \in \mathbb{I}^2$  con  $X$  sorteado uniformemente en  $\mathbb{I}^2$  y  $\mathfrak{D}_k$  una partición según el algoritmo antes descrito. Entonces  $\text{diam}(A(x)) \xrightarrow{k \rightarrow +\infty} 0$  en probabilidad.*

*Demostración.* Por conveniencia, denotemos por  $Z$  al vector aleatorio  $k$ -dimensional que determina la partición  $\mathfrak{D}_k$  (cada entrada es un vector aleatorio  $(r_i, c_i, U_i)$ , donde  $r_i$  representa la celda seleccionada,  $c_i$  es una bernoulli(1/2) representando la coordenada, y  $U_i$  es una uniforme  $U(0, 1)$  que representa en qué proporción partimos el lado del rectángulo, todas independientes). Debemos probar que  $\text{diam}(A(X, Z)) \rightarrow 0$  en probabilidad. Para eso, nótese como  $V(X, Z)$  el tamaño del lado correspondiente la primera coordenada de la celda  $A(X)$ . Sea  $T(X, Z)$  el número de veces que  $A(x)$  es partida en el proceso de construcción del árbol aleatorio. Sea  $K$  una binomial  $(T, 1/2)$ , que representa la cantidad de veces que se parte la celda que contiene a  $X$  por la primera coordenada.

Con esta notación, basta probar que  $V(X, Z) \rightarrow 0$  en probabilidad, ya que el mismo argumento que se hace para la primera coordenada, se podrá hacer para el resto. Como  $V(X, Z) \geq 0$ , y a través de la desigualdad de Markov, basta probar que  $\mathbb{E}_{X, Z} [V(X, Z)] \rightarrow 0$ . Fijado  $X = x$  y tomando esperanza en  $Z$  (o sea, tomando esperanza en  $\mathfrak{D}_k$ ), tenemos:

$$\begin{aligned}
\mathbb{E}_Z [V(x, Z)] &\leq \mathbb{E}_K \left[ \mathbb{E}_{U_i} \left[ \prod_{i=1}^K \text{máx}(U_i, 1 - U_i) \mid K_n \right] \right] \\
&= \mathbb{E}_K \left[ (\mathbb{E}_{U_1} [\text{máx}(U_1, 1 - U_1)])^K \right] \\
&= \mathbb{E}_K \left[ (3/4)^K \right] = \mathbb{E}_T \left[ \mathbb{E}_K \left[ (3/4)^K / T \right] \right] \\
&= \mathbb{E}_T \left[ \sum_{i=0}^T \frac{3^i}{4} \binom{T}{i} \frac{1^i}{2} \left(1 - \frac{1}{2}\right)^{T-i} \right] \\
&= \mathbb{E}_T \left[ \sum_{i=0}^T \binom{T}{i} \frac{3^i}{8} \left(1 - \frac{1}{2}\right)^{T-i} \right] \\
&= \mathbb{E}_T \left[ \left(1 - \frac{1}{2} + \frac{3}{8}\right)^T \right] \\
&= \mathbb{E}_T \left[ \left(1 - \frac{1}{8}\right)^T \right]
\end{aligned}$$

En la primera desigualdad, se descompone la aleatoriedad del proceso de construcción del árbol aleatorio ( $Z$ ). Como sólo interesa la primera coordenada, se fija cuántas veces corta la primera coordenada ( $K$ ), y a qué altura va cortando ( $U_i$ ). El largo del lado de la primera coordenada de la celda que contiene a  $x$ , luego del primer corte por dicho lado, estará acotada por el máximo de  $U_1$  y  $1 - U_1$ . Pensando análogamente para los  $K$  cortes, se obtiene la desigualdad. Luego se usa que, al ser  $U_i$  independientes e idénticamente distribuidos, el producto de las esperanzas es la esperanza del producto. El siguiente paso se justifica porque el  $\text{máx}(U_1, 1 - U_1) \sim U[1/2, 1]$  entonces  $\mathbb{E}[\text{máx}(U_1, 1 - U_1)] = 3/4$ . Posteriormente como  $K \sim \text{Binomial}(T, \frac{1}{2})$  y  $K = \{0, 1, 2 \dots T\}$  se calcula  $\mathbb{E}_K [(3/4)^K]$ . Finalmente por el binomio de Newton se llega a la última igualdad.

El problema se resume en probar que  $T \rightarrow \infty$ . Como  $Z$  nos devuelve un árbol binario, se usa el teorema que dice que, si  $k$  es grande, todos los niveles del árbol hasta  $\ell = \lfloor 0, 37 \rfloor \log(k)$  están llenos con probabilidad cercana a 1 ([7]). Esto asegura que todas las celdas fueron partidas más de  $\ell$  veces con probabilidad cercana a 1, y por ende  $T \rightarrow \infty$  en probabilidad.

□

## 2.3. Convergencia

Tenemos entonces una muestra  $\mathbb{M}_n \subset \mathbf{I}^2$  proveniente de una cópula  $C$  con densidad  $p$ . El objetivo ahora es probar que  $p_k$  (definido en 2.2) converge a  $p$ , con  $k$  tendiendo a infinito. Sin embargo,  $k$  no puede crecer demasiado en comparación a como crece el tamaño de la muestra, es importante que cada celda tenga “infinitas” observaciones. Lo que haremos a continuación será probar que el tree estimator  $\hat{p}$  (ver 2.2) converge, y la convergencia de  $p_k$  será un corolario. La idea de la demostración está basada en el teorema 6.1 de [6] y en el teorema 3.1 de [2].

**Notación.** Notaremos  $\mathbb{E}_\lambda(f) = \int_{\mathcal{I}^2} f(x, y) dx dy$ .

**Teorema 2.3.1.** *Sea  $\mathbb{X}_n = X_1, \dots, X_n$  una muestra aleatoria simple, con  $X_i \in \mathbf{I}^2$ , proveniente de una distribución bivariada, soportada en  $\mathbf{I}^2$ , con marginales uniformes  $U(0, 1)$  y densidad  $p$ . Sea  $\mathfrak{D}_k$  una partición aleatoria de  $\mathbf{I}^2$  (definida en la sección anterior) y  $\lambda$  la medida de Lebesgue. Entonces se cumple*

$$\mathbb{E}_{\lambda_{\mathbb{X}_n \mathfrak{D}_k}}(|\hat{p}(x) - p(x)|) \rightarrow 0$$

cuando  $n \rightarrow +\infty$ ,  $k \rightarrow +\infty$  y además  $k/\sqrt{n} \rightarrow 0$ .

*Demostración.* Primero recordemos la notación. Sean  $A_1, \dots, A_{k+1}$  los subrectángulos (celdas) que conforman  $\mathfrak{D}_k$ , y denotemos por  $A(x)$  a la celda que contiene a  $x$ . Sea  $N(x) = \#\{X_i | X_i \in A\}$  la cantidad de observaciones que están en la celda que contiene a  $x$ , y  $N(A_i)$  la cantidad de observaciones en la celda  $A_i$ . Tenemos que el tree estimator tiene la forma

$$\hat{p}(x) = \frac{N(x)}{n\lambda(A(x))}.$$

La “versión teórica” del tree estimator  $\hat{p}(x)$  se define de la siguiente manera:

$$\begin{aligned} \bar{p}(x) = \mathbb{E}(p(x) | x \in A(x)) &= \frac{\int_{A(x)} p(x) d\lambda(x)}{\lambda(A(x))} \\ &= \frac{\mathbb{P}_X(A(x))}{\lambda(A(x))} \end{aligned}$$

Intuitivamente, si tuvieramos una cantidad de observaciones tendiendo a infinito, y una partición fija,  $\hat{p} \rightarrow \bar{p}$  por la ley fuerte de los grandes números. Por la desigualdad triangular tenemos:

$$\mathbb{E}(|p(x) - \hat{p}(x)|) \leq \mathbb{E}(|\bar{p}(x) - \hat{p}(x)|) + \mathbb{E}(|p(x) - \bar{p}(x)|)$$

y la idea será demostrar que ambos términos tienden a cero.



Estudiamos el primer término:  $\mathbb{E}(|\bar{p}(x) - \hat{p}(x)|)$ .

Tomando la esperanza sobre  $\lambda$  (o sea, fijando  $\mathfrak{D}_k$  y la MAS) obtenemos:

$$\begin{aligned}\mathbb{E}_\lambda(|\bar{p}(x) - \hat{p}(x)|) &= \sum_{i=1}^{k+1} \left| \frac{\mathbb{P}_X(A_i)}{\lambda(A_i)} - \frac{N(A_i)}{n\lambda(A_i)} \right| \lambda(A_i) \\ &= \sum_{i=1}^{k+1} \left| \mathbb{P}_X(A_i) - \frac{N(A_i)}{n} \right|\end{aligned}$$

Ahora queremos tomar esperanza sobre  $\mathbb{X}_n$  (manteniendo fija la partición  $\mathfrak{D}_k$  y el tamaño de la muestra  $n$ ). Para eso, fijemos una celda  $A(x)$ , y observemos que la probabilidad que suceda  $N(x) = \alpha$  puede expresarse como una binomial  $Bin_{n, \mathbb{P}_X(A(x))}(\alpha)$ . Entonces

$$\begin{aligned}\mathbb{E} \left( \sum_{i=1}^{k+1} \left| \mathbb{P}_X(A_i) - \frac{N(A_i)}{n} \right| \right) &= \sum_{i=1}^{k+1} \mathbb{E} \left( \left| \mathbb{P}_X(A_i) - \frac{N(A_i)}{n} \right| \right) \\ &= \frac{1}{n} \sum_{i=1}^{k+1} \mathbb{E} (|n\mathbb{P}_X(A_i) - Bin_{n, \mathbb{P}_X(A_i)}|) \\ &\leq \frac{1}{n} \sum_{i=1}^{k+1} \sqrt{Var(Bin(n, \mathbb{P}_X(A_i)))} \\ &= \frac{1}{n} \sum_{i=1}^{k+1} \sqrt{n\mathbb{P}_X(A_i)(1 - \mathbb{P}_X(A_i))} \\ &= \frac{k+1}{2\sqrt{n}} \rightarrow 0\end{aligned}$$

La desigualdad es consecuencia de la desigualdad de Cauchy-Schwarz, y la última igualdad surge de acotar  $\mathbb{P}_X(A_i)(1 - \mathbb{P}_X(A_i)) \leq 1/4$ . Como la cota resultante no depende de  $\mathfrak{D}_k$  (solo depende de  $k$ ) entonces podemos concluir que  $\mathbb{E}(|\bar{p}(x) - \hat{p}(x)|) \rightarrow 0$ .

Veamos ahora el segundo término:  $\mathbb{E}(|p(x) - \bar{p}(x)|)$ .

La idea será fijar  $\epsilon$  arbitrariamente pequeño, y demostrar que para  $k$  suficientemente grande se tiene  $\mathbb{E}(|p(x) - \bar{p}(x)|) \leq \epsilon$  (c.s). Sea  $p_\delta$  una función continua y acotada (con cota  $M$ ), definida en  $\mathbf{I}^2$ , tal que  $p_\delta|_{\partial\mathbf{I}^2} = 0$  y  $\mathbb{E}|p(x) - p_\delta| \leq \delta$  (tomando la esperanza respecto de  $\lambda$ ). Dicha función existe por ser las funciones continuas y acotadas densas en  $L^1(\mathbf{I}^2)$ , y multiplicando por una “función chichón” (lema de Urysohn) logramos la condición de borde.

La condición de borde se pide por si no podemos extender  $p$  a  $\mathbf{I}^2$  (como en los casos donde hay tail dependence y  $p$  tiende a infinito cerca del origen).

De manera análoga tenemos

$$\bar{p}_\delta(x) = \frac{\int_{A(x)} p_\delta d\lambda}{\lambda(A(x))}.$$

De esta manera podemos afirmar:

$$\begin{aligned} \mathbb{E}(|p(x) - \bar{p}(x)|) &\leq \mathbb{E}(|p(x) - p_\delta(x)|) + \\ &\quad + \mathbb{E}(|p_\delta(x) - \bar{p}_\delta(x)|) + \\ &\quad + \mathbb{E}(|\bar{p}_\delta(x) - \bar{p}(x)|) \end{aligned}$$

Tomemos  $\delta \leq \epsilon/3$ . El primer término es menor a  $\delta$  por definición de  $p_\delta$ . Con respecto al tercer término tenemos

$$\begin{aligned} \delta &\geq \mathbb{E}(|p(x) - p_\delta(x)|) \\ &= \sum_{i=1}^{k+1} \lambda(A_i) \mathbb{E} \left| \frac{(p(x) - p_\delta(x)) \mathbf{1}_{A_i}}{\lambda(A_i)} \right| \\ &\geq \mathbb{E}|\bar{p}(x) - \bar{p}_\delta| \end{aligned}$$

donde la última desigualdad surge de la desigualdad de Jensen, y de que  $\bar{p}(x) - \bar{p}_\delta$  es constante en cada celda.

Basta acotar el segundo término. Sabemos que  $p_\delta$  es continua en un dominio compacto, y por tanto es uniformemente continua. Por lo tanto, fijado  $\mu$  arbitrariamente chico, existe  $\eta$  tal que si la celda  $A_i$  tiene diámetro menor a  $\eta$ , entonces

$$\int_{A_i} |p_\delta - \bar{p}_\delta| d\lambda \leq \mu \lambda(A_i).$$

Sea  $\mu = \delta$ . Si denotamos por  $B$  al suceso “todas las celdas tienen diámetro menor a  $\eta$ ”, sabemos que  $\mathbb{P}_{\mathfrak{D}_k}\{B\} \xrightarrow{k \rightarrow +\infty} 1$ . Por lo tanto, tomando esperanza sobre  $\lambda$  y sobre  $\mathfrak{D}_k$  (no depende de la MAS), obtenemos:

$$\begin{aligned} \mathbb{E}(|p_\delta(x) - \bar{p}_\delta(x)|) &= \mathbb{E}(|p_\delta(x) - \bar{p}_\delta(x)| \mathbf{1}_B) + \mathbb{E}(|p_\delta(x) - \bar{p}_\delta(x)| (1 - \mathbf{1}_B)) \\ &\leq \delta \mathbb{P}_{\mathfrak{D}_k}\{B\} + 2M(1 - \mathbb{P}_{\mathfrak{D}_k}\{B\}) \xrightarrow{k \rightarrow +\infty} \delta \end{aligned}$$

Por último, teniendo los tres términos acotados, tenemos que dado  $k$  suficientemente grande,  $\mathbb{E}(|p(x) - \bar{p}(x)|) \leq 3\delta = \epsilon$ , lo cual concluye la demostración. □

Dos rápidos corolarios surgen del teorema 2.3.1.

**Corolario.** *En las hipótesis del teorema anterior, tenemos*

$$\mathbb{E}_{\lambda_{\mathbb{X}_n}}(|\mathbb{E}_{\mathfrak{D}_k} \hat{p}(x) - p(x)|) = \mathbb{E}_{\lambda_{\mathbb{X}_n}}(|p_k(x) - p(x)|) \rightarrow 0$$

con  $n \rightarrow +\infty$ ,  $k \rightarrow +\infty$  y  $k/\sqrt{n} \rightarrow 0$ .

**Corolario.** *En las hipótesis del teorema anterior, si para un mismo  $k$  y una muestra  $\mathbb{X}_n$  sorteamos  $m$  distintos tree estimators  $\hat{p}_1, \dots, \hat{p}_m$ , entonces se tiene*

$$\mathbb{E}_{\lambda_{\mathbb{X}_n} \mathfrak{D}_k^{1, \dots, m}} \left( \left| \frac{1}{m} \sum_{i=1}^m \hat{p}_i(x) - p(x) \right| \right) \rightarrow 0$$

con  $n \rightarrow +\infty$ ,  $k \rightarrow +\infty$  y  $k/\sqrt{n} \rightarrow 0$ .

En el teorema 2.3.1, definimos la versión teórica del tree estimator  $\hat{p}$  como

$$\begin{aligned} \bar{p}(x) = \mathbb{E}(p(x)|x \in A(x)) &= \frac{\int_{A(x)} p(x) d\lambda(x)}{\lambda(A(x))} \\ &= \frac{\mathbb{P}_X(A(x))}{\lambda(A(x))} \end{aligned}$$

y de la misma manera, podemos definir la versión teórica del random forest estimator  $p_k$  como

$$\bar{p}_k(x) = \mathbb{E}_{\mathfrak{D}_k}(\bar{p}(x))$$

*Observación 2.3.2.* En la demostración 2.3.1, cuando demostramos que el término  $\mathbb{E}(|\bar{p}(x) - \hat{p}(x)|)$  tiende a cero, solo necesitamos que  $k/\sqrt{n} \rightarrow 0$ , y eso sucede tomando  $k$  fijo y  $n \rightarrow +\infty$ . Por lo tanto, se tiene que para un  $k$  fijo y  $n \rightarrow +\infty$ , el random forest estimator  $p_k$  converge a  $\bar{p}_k$ .

Obviamente, nunca vamos a tener una muestra aleatoria simple infinita, así que siempre nos tendremos que conformar con una muestra de tamaño “grande”. Para entender mejor como funciona el random forest estimator para cierto tamaño de muestra  $n$  y para cierta cantidad de cortes  $k$  (con  $n \gg k$ ), calculemos  $\bar{p}_k(x)$  para  $k = 2$ .

**Proposición 2.3.3.** *Sea  $p$  como en las hipótesis del teorema 2.3.1. Sea  $(x, y) \in \mathcal{I}^2$ . Entonces*

1.  $\bar{p}_1(x, y) = 1$ .

2.

$$\begin{aligned} \bar{p}_2(x, y) = & \frac{3}{4} + \frac{1}{4} \left[ \int_x^1 \int_y^1 \left( \frac{1}{\alpha\beta} \int_0^\alpha \int_0^\beta p(a, b) dbda \right) d\beta d\alpha \right. \\ & + \int_0^x \int_y^1 \left( \frac{1}{(1-\alpha)\beta} \int_\alpha^1 \int_0^\beta p(a, b) dbda \right) d\beta d\alpha \\ & + \int_x^1 \int_0^y \left( \frac{1}{\alpha(1-\beta)} \int_0^\alpha \int_\beta^1 p(a, b) dbda \right) d\beta d\alpha \\ & \left. + \int_0^x \int_0^y \left( \frac{1}{(1-\alpha)(1-\beta)} \int_\alpha^1 \int_\beta^1 p(a, b) dbda \right) d\beta d\alpha \right] \end{aligned}$$

*Demostración.*

1. Sea  $(x, y) \in \mathbf{I}^2$ . Mostraremos que  $\bar{p}_1(x, y) = 1$  observando que si  $k = 1$  entonces  $\bar{p}(x, y) = 1$ .

Para ello, supongamos sin pérdida de generalidad que el corte se hace perpendicular al eje  $\bar{x}$  a través de  $u \in \mathbf{I}$ , y  $u > x$ . Entonces

$$\bar{p}(x, y) = \frac{1}{\lambda(A(x, y))} \int_0^u \int_0^1 p(a, b) dadb = \frac{u}{u} = 1$$

donde usamos que las distribuciones marginales son uniformes  $U(0, 1)$ .

2. Sea  $(x, y) \in \mathbf{I}^2$ . A continuación calcularemos  $\bar{p}_2(x, y)$  condicionando, discutiendo según las distintas particiones de  $\mathbf{I}^2$  de dos cortes, y a cual de los rectángulos de la partición pertenece  $(x, y)$ . La primera discusión es si ambos cortes se realizan sobre la misma coordenada, o sobre coordenadas distintas (ambas opciones tienen probabilidad  $1/2$ ). Si hay cortes paralelos, se tiene  $\bar{p}(x, y) = 1$  por las mismas ideas del estimador de un corte. En caso de cortes perpendiculares, la discusión es sobre si  $(x, y)$  pertenece a la celda que es cortada 2 veces o no (ambas opciones tienen probabilidad  $1/2$ ). Si la celda solo fue cortada una vez, de vuelta  $\bar{p}(x, y) = 1$ . La última discusión es sobre las cuatro posibilidades de la celda que contiene a  $(x, y)$ , ya que uno de los vértices de la celda tiene que ser el  $(0, 0)$ , o el  $(1, 0)$ , o el  $(0, 1)$  o el  $(1, 1)$ . Por ejemplo, si uno de los vértices es el  $(0, 0)$ , el vértice opuesto  $(\alpha, \beta)$  tiene que cumplir  $x < \alpha$ ,  $y < \beta$ ,  $\alpha$  es equiprobable en

$(x, 1)$ , y  $\beta$  es equiprobable en  $(y, 1)$ . A partir de eso podemos tomar esperanza sobre todos los posibles  $\alpha$  y  $\beta$ .

$$\begin{aligned}
\bar{p}_2(x, y) &= \frac{1}{2}(\text{cortes paralelos}) + \frac{1}{2}(\text{cortes perpendiculares}) \\
&= \frac{1}{2} + \frac{1}{2} \left[ \frac{1}{2}((x, y) \text{ en celda grande}) + \frac{1}{2}((x, y) \text{ en celda chica}) \right] \\
&= \frac{3}{4} + \frac{1}{4} \left[ \int_x^1 \int_y^1 \left( \frac{1}{\alpha\beta} \int_0^\alpha \int_0^\beta p(a, b) dbda \right) d\beta d\alpha \right. \\
&\quad + \int_0^x \int_y^1 \left( \frac{1}{(1-\alpha)\beta} \int_\alpha^1 \int_0^\beta p(a, b) dbda \right) d\beta d\alpha \\
&\quad + \int_x^1 \int_0^y \left( \frac{1}{\alpha(1-\beta)} \int_0^\alpha \int_\beta^1 p(a, b) dbda \right) d\beta d\alpha \\
&\quad \left. + \int_0^x \int_0^y \left( \frac{1}{(1-\alpha)(1-\beta)} \int_\alpha^1 \int_\beta^1 p(a, b) dbda \right) d\beta d\alpha \right]
\end{aligned}$$

□

*Observación 2.3.4.* Recordemos que  $\bar{p}_k$  es una densidad, por lo tanto  $\int_{\mathbf{I}^2} \bar{p}_2(a, b) dbda = 1$ , y además el segundo término de  $\bar{p}_2$  es positivo, por lo tanto  $\bar{p}_2(x, y) \geq \frac{3}{4}$ . Por lo tanto, podemos pensar que  $\bar{p}_2$  (y  $\bar{p}_k$  en general) suaviza y contrae la densidad  $p$  hacia la función constante 1.

## 2.4. Aplicaciones numéricas

Para probar nuestro programa, simulamos una muestra aleatoria simple de 500 datos utilizando el paquete “copula” del software R (el tamaño de muestra aproximado de los datos que nos interesarán). Por ejemplo, simulamos datos provenientes de una cópula Gumbel utilizando el paquete “copula” del software **R** (más detalles sobre la cópula Gumbel en el siguiente capítulo). Probamos esta densidad porque presenta dificultades que se pueden presentar en nuestro problema 4.1.4; por ejemplo, dependencia de colas superior distinta a 0 (lo cual implica que la densidad tiende a  $+\infty$  en  $(1, 1)$ ), y tiene un pequeño máximo local en  $(0, 0)$ . Nos interesará que el estimador identifique ambos máximos locales, y que el máximo local en  $(1, 1)$  sea mayor al máximo local en  $(0, 0)$ .

En estas simulaciones, el error en norma  $L^1$  (calculado numéricamente) entre la densidad de la cópula Gumbel con  $\theta = 3$  y el estimador por random forest es aproximadamente 0,69 (observamos que este error siempre estará comprendido entre 0 y 2).

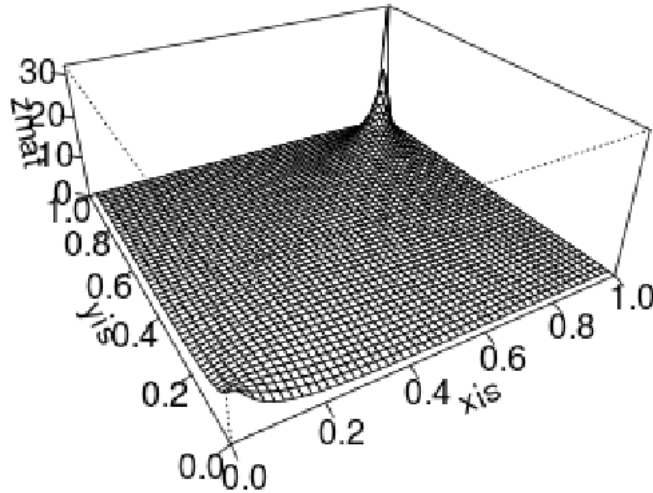


Figura 2.1: Densidad de la cópula Gumbel de parámetro  $\theta = 3$

Algunas observaciones que se pueden realizar en base a estas estimaciones:

- Nuestro estimador parece ser robusto; esto es esperable dado el azar y el promediado que hay en el programa.
- Nuestro estimador parece respetar la forma de la densidad original. Podemos pensar que nuestro estimador puede usarse para conjeturar con qué familias (paramétricas) podemos obtener buenos ajustes, ya que saber dónde hay máximos y mínimos locales y globales en la densidad ayuda a descartar algunos modelos.
- Nuestro estimador subestima máximos y sobrestima mínimos. Sobre esta observación, recordemos 2.3.4, y además pensemos que si discretizamos el espacio en el programa, los grandes valores (que se dan en pequeñas áreas de  $\mathbf{I}^2$ ) se distribuyen en mayores áreas. Concluimos que no es buena idea utilizar este estimador, por ejemplo, para estimar dependencia de colas.
- Nuestro estimador parece ser, como mínimo, competente frente a la estimación por kernel. La estimación por kernel presenta defectos, en el caso de estimación de densidades de cópulas, en términos de la

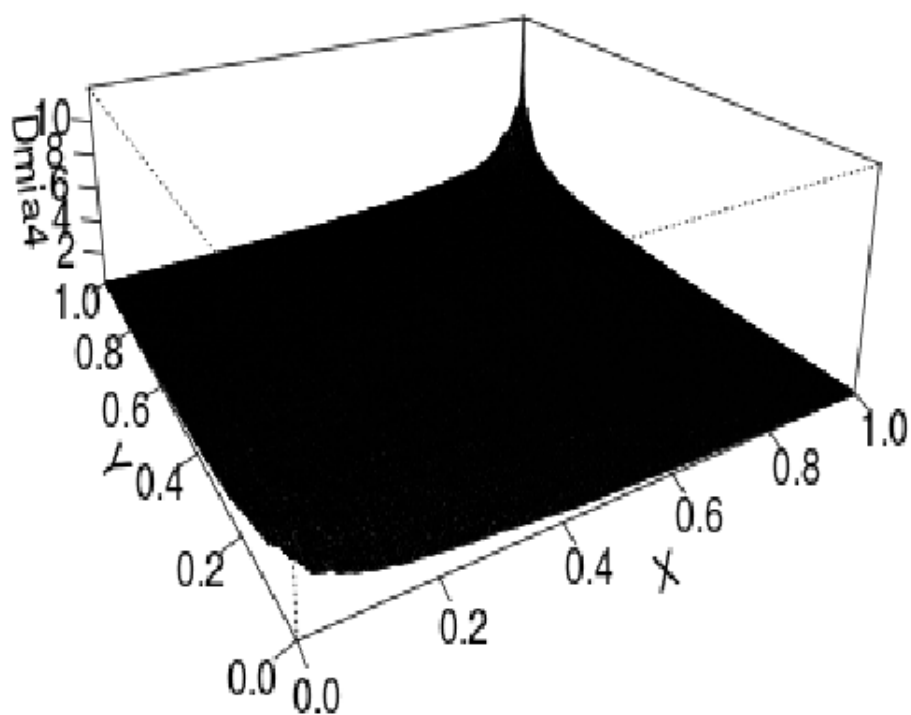


Figura 2.2: Estimación de la densidad de la cópula Gumbel de parámetro  $\theta = 3$ .  $n = 500$  datos,  $k = 100$ , tamaño de la grilla base  $g = 450$ , cantidad de árboles  $t = 2000$ .

compacidad del soporte, y de la invariancia en la rotación de los ejes (propiedad deseable en estimaciones de densidades en general, pero causa efectos indeseados cuando los ejes tienen particular interpretación). Basados en simulaciones, observamos que los resultados de la estimación por kernel de la cópula Gumbel son especialmente pobres cuando la muestra tiene menos de 1500 datos.



## Capítulo 3

# Estimación paramétrica

En este capítulo presentaremos herramientas para realizar estimaciones paramétricas de densidades de cópulas. A diferencia de las estimaciones no paramétricas de densidades de cópulas, primero debemos proponer una familia de cópulas  $\mathfrak{F}_\theta = \{c_\theta | \theta \in \Theta\}$ , y supondremos que la densidad  $c$  de la cual proviene nuestra muestra pertenece a  $\mathfrak{F}_\theta$ . Una vez fijada  $\mathfrak{F}_\theta$ , primero debemos estimar el parámetro  $\hat{\theta}$  que hace que  $c_{\hat{\theta}}$  logre el mejor ajuste a los datos dentro de  $\mathfrak{F}_\theta$  (en nuestro caso máxima verosimilitud); debemos tener en cuenta que aunque el ajuste sea el mejor dentro de la familia, no necesariamente es un buen ajuste. Luego estudiaremos si, efectivamente,  $c_{\hat{\theta}}$  es un buen candidato como aproximación de  $c$  (aplicaremos test chi cuadrado para bondad de ajuste).

Nosotros nos concentraremos en estimaciones paramétricas con familias de cópulas que estén contenidas dentro de las *cópulas arquimedianas*. Si bien las cópulas arquimedianas presentan cierta rigidez (son simétricas respecto a la diagonal principal, por ejemplo), pueden cubrir una variedad muy amplia de subfamilias (distintos tipos de dependencia de colas por ejemplo, cosa que nos viene bien). Además, las cópulas arquimedianas presentan propiedades muy deseables (se construyen a partir de una función generatriz, por ejemplo), que hacen que los cálculos se faciliten.

La primera sección la dedicaremos a la presentación de las familias arquimedianas y sus propiedades básicas; además, presentaremos las cuatro familias que utilizaremos para la estimación paramétrica: con ellas abarcaremos una gran variedad de comportamientos de los datos. En la segunda y tercera sección haremos una breve reseña de la estimación por máxima verosimilitud y del test chi cuadrado respectivamente.

### 3.1. Familias arquimedianas

Nos basamos en el capítulo 4 de [1].

Nos disponemos ahora a encontrar cópulas que cumplan

$$\varphi(C(u, v)) = \varphi(u) + \varphi(v).$$

La función  $\varphi$  se llamará **función generatriz** de la cópula  $C$

A nosotros nos interesará poder escribir  $C$  en función de  $\varphi$ , por lo tanto necesitamos una noción de inversa para las funciones generatrices.

**Definición 3.1.1.** Sea  $\varphi: \mathbf{I} \rightarrow [0, +\infty]$  una función continua, estrictamente decreciente, tal que  $\varphi(1) = 0$ . La **pseudo-inversa** de  $\varphi$  es la función  $\varphi^{[-1]}: [0, +\infty] \rightarrow \mathbf{I}$  dada por

$$\varphi^{[-1]}(t) = \begin{cases} \varphi^{-1}(t) & \text{si } 0 \leq t \leq \varphi(0) \\ 0 & \text{en otro caso} \end{cases}$$

*Observación 3.1.2.*

$$\begin{aligned} \varphi\left(\varphi^{[-1]}(t)\right) &= \begin{cases} t & \text{si } 0 \leq t \leq \varphi(0) \\ \varphi(0) & \text{en otro caso} \end{cases} \\ &= \min\{t, \varphi(0)\} \end{aligned}$$

Además, si  $\varphi(0) = +\infty$ , entonces  $\varphi^{[-1]} = \varphi^{-1}$ .

A partir de lo anterior, podemos escribir

$$C(u, v) = \varphi^{[-1]}(\varphi(u) + \varphi(v)),$$

por lo cual la idea es estudiar algunas condiciones necesarias que debe satisfacer  $\varphi$  para que  $C$  sea una cópula.

**Proposición 3.1.3.** Sea  $\varphi: \mathbf{I} \rightarrow [0, +\infty]$  una función continua, estrictamente decreciente, tal que  $\varphi(1) = 0$ . Entonces la función  $C: \mathbf{I}^2 \rightarrow \mathbf{I}$  definida por

$$C(u, v) = \varphi^{[-1]}(\varphi(u) + \varphi(v))$$

cumple las condiciones de borde de una cópula.

*Demostración.*    ■  $C(0, v) = \varphi^{[-1]}(\varphi(0) + \varphi(v)) = 0$ .

■  $C(1, v) = \varphi^{[-1]}(\varphi(1) + \varphi(v)) = v$ ,

Donde en la primera igualdad usamos  $\varphi(0) + \varphi(v) \geq \varphi(0)$  junto con la definición de  $\varphi^{[-1]}$ , y en la segunda igualdad usamos  $\varphi(1) = 0$ .    □

Más aún, si  $\varphi$  cumple las hipótesis de la proposición 3.1.3, podemos pedir una hipótesis adicional, que será suficiente y necesaria para que  $C$  sea una cópula ([1], página 111, teorema 4.1.4).

**Teorema 3.1.4.** *Supongamos que las hipótesis de la proposición 3.1.3 se cumplen. Entonces  $C$  es una cópula si y solo si  $\varphi$  es convexa.*

Llamaremos **cópulas arquimedianas** a las cópulas que podemos generar con el teorema 3.1.4.

**Definición 3.1.5.** Una función  $\varphi$  que cumpla las hipótesis del teorema 3.1.4 se llama **función generatriz** de la cópula arquimediana  $C$ . Si además  $\varphi(0) = +\infty$ , entonces decimos que  $\varphi$  es una **función generatriz estricta**, y que la cópula arquimediana  $C$  es estricta.

Algunas propiedades interesantes de las cópulas arquimedianas son las siguientes.

**Proposición 3.1.6.** *Sea  $C$  una cópula arquimediana con función generatriz  $\varphi$ . Entonces*

- $C$  es simétrica:  $C(u, v) = C(v, u)$ .
- $C$  es asociativa:  $C(C(u, v), w) = C(u, C(v, w))$
- Si  $c > 0$ , entonces  $c\varphi$  también es una función generatriz de  $C$

*Demostración.*    ■  $C(u, v) = \varphi^{[-1]}(\varphi(u) + \varphi(v)) = C(v, u)$ .

■ Tenemos:

$$\begin{aligned} C(C(u, v), w) &= \varphi^{[-1]} \left( \varphi \left( \varphi^{[-1]}(\varphi(u) + \varphi(v)) \right) + \varphi(w) \right) \\ &= \varphi^{[-1]} \left( \min\{\varphi(u) + \varphi(v), \varphi(0)\} + \varphi(w) \right) \\ &= \varphi^{[-1]} \left( \min\{\varphi(u) + \varphi(v) + \varphi(w), \varphi(0) + \varphi(w)\} \right) \\ &= \varphi^{[-1]} \left( \min\{\varphi(u) + \varphi(v) + \varphi(w), \varphi(0)\} \right). \end{aligned}$$

Donde en la última utilizamos la definición de función pseudo-inversa, y que  $\varphi(w) \geq 0$ . De manera análoga se llega a que

$$C(u, C(v, w)) = \varphi^{[-1]}(\min\{\varphi(u) + \varphi(v) + \varphi(w), \varphi(0)\}),$$

con lo cual queda demostrado.

- Definamos  $\psi = c\varphi$ . Lo primero, es que como  $\varphi$  es una función generatriz, entonces cumple las hipótesis del teorema 3.1.4. Claramente si  $c > 0$  entonces  $\psi$  también cumplirá las hipótesis del teorema 3.1.4, y existirá una cópula arquimediana  $\hat{C}$  de la cual  $\psi$  es función generatriz. Se tendrá que  $\hat{C} = C$  si  $C(u, v) = \psi^{[-1]}(\psi(u) + \psi(v))$ . En efecto:

$$\begin{aligned} C(u, v) = \psi^{[-1]}(\psi(u) + \psi(v)) &\iff \\ \psi(C(u, v)) = \psi(u) + \psi(v) &\iff \\ c\varphi(C(u, v)) = c\varphi(u) + c\varphi(v) & \end{aligned}$$

y la última igualdad es verdadera por ser  $\varphi$  una función generatriz de  $C$ . □

Una propiedad interesante de las cópulas arquimedias es que, dada una función generatriz  $\varphi$ , podemos generar una familia de funciones generatrices de dos parámetros  $\alpha, \beta$ . Recordamos antes un lema previo.

**Lema 3.1.7.** *Sea  $f: A \rightarrow \mathbb{R}$  una función convexa, y  $g: B \rightarrow \mathbb{R}$  una función convexa y no decreciente, donde  $A$  y  $B$  son subconjuntos conexos de  $\mathbb{R}$  tales que  $f(A) \subset B$ . Entonces  $g \circ f$  es convexa.*

*Demostración.* Como  $f$  es convexa, tenemos que si  $\lambda \in \mathbf{I}$ , entonces  $f(\lambda u + (1 - \lambda)v) \leq \lambda f(u) + (1 - \lambda)f(v)$ . Usando primero que  $g$  es creciente y luego que es convexa, tenemos

$$\begin{aligned} g(f(\lambda u + (1 - \lambda)v)) &\leq g(\lambda f(u) + (1 - \lambda)f(v)) \\ &\leq \lambda g(f(u)) + (1 - \lambda)g(f(v)) \end{aligned}$$

de lo cual concluimos que  $g \circ f$  es convexa. □

**Proposición 3.1.8.** *Sea  $\varphi(t)$  una función generatriz de una cópula arquimediana. Entonces*

1.  $\varphi_{\alpha,1}(t) = \varphi(t^\alpha)$ , con  $\alpha \in (0, 1]$ , es una función generatriz de una cópula arquimediana.

2.  $\varphi_{1,\beta}(t) = (\varphi(t))^\beta$ , con  $\beta \in [1, +\infty)$ , es una función generatriz de una cópula arquimediana.

Concluimos que  $\varphi_{\alpha,\beta}(t) = (\varphi(t^\alpha))^\beta$ , con  $\alpha$  y  $\beta$  como antes, es una función generatriz de una cópula arquimediana.

*Demostración.* La demostración se basa en el lema 3.1.7, componiendo convenientemente.

1. Tomando  $f(t) = -t^\alpha$  y  $g(t) = \varphi(-t)$ , se tiene que  $f$  y  $g$  entran en las hipótesis del lema 3.1.7, y  $g \circ f = \varphi_{\alpha,1}$ .
2. Tomando  $f(t) = \varphi(t)$  y  $g(t) = t^\beta$ , se tiene que  $f$  y  $g$  entran en las hipótesis del lema 3.1.7, y  $g \circ f = \varphi_{1,\beta}$ .

□

*Observación 3.1.9.* Las hipótesis de la proposición 3.1.8 son suficientes pero no necesarias. Esto es, puede suceder que podamos extender el conjunto de los posibles  $\alpha$  y  $\beta$ , dependerá de la función generatriz  $\varphi$  que consideremos.

Más aún, podemos expresar la dependencia de colas de  $\varphi_{\alpha,1}$  y la de  $\varphi_{1,\beta}$  en función de la dependencia de colas de  $\varphi$  (ver [3]).

**Teorema 3.1.10.** *Sea  $\varphi$  una función generatriz de una cópula arquimediana, con dependencia de colas superior  $\lambda_U$  y dependencia de colas inferior  $\lambda_L$ . Consideremos las funciones generatrices  $\varphi_{\alpha,1}$  y  $\varphi_{1,\beta}$ . Llamemos  $\lambda_U^{(\alpha)}$  y  $\lambda_L^{(\alpha)}$  a las dependencias de colas superior e inferior respectivamente de la cópula generada por  $\varphi_{\alpha,1}$ , y análogamente llamemos  $\lambda_U^{(\beta)}$  y  $\lambda_L^{(\beta)}$  a las dependencias de colas superior e inferior respectivamente de la cópula generada por  $\varphi_{1,\beta}$ . Entonces*

1.  $\lambda_U^{(\alpha)} = \lambda_U$ ,  $\lambda_L^{(\alpha)} = (\lambda_L)^{1/\alpha}$ .
2.  $\lambda_U^{(\beta)} = 2 - (2 - \lambda_U)^{1/\beta}$ ,  $\lambda_L^{(\beta)} = \lambda_L^{1/\beta}$ .

Más adelante consideraremos la función generatriz  $\varphi(t) = (1/t) - 1$ , y la familia de dos parámetros  $\varphi_{\alpha,\beta}$  generada por la proposición anterior.

Otra propiedad muy útil de las cópulas arquimedianas es que, si  $C$  es una cópula arquimediana con función generatriz  $\varphi$ , se puede escribir la densidad de  $C$  (si existe) en función de  $C$  y las derivadas primera y segunda de  $\varphi$ .

$$c(u, v) = \frac{\partial^2 C}{\partial u \partial v}(u, v) = -\frac{\varphi''(C(u, v))\varphi'(u)\varphi'(v)}{[\varphi'(C(u, v))]^3}.$$

Esto es particularmente útil a la hora de estimar el parámetro de ajuste via máxima verosimilitud en un modelo paramétrico.

Si siguiendo la misma línea, podemos expresar la dependencia de colas de una cópula arquimediana  $C$  con función generatriz  $\varphi$  en función de la pseudo-inversa  $\varphi^{[-1]}$ .

$$\lambda_L = \lim_{t \rightarrow 0^+} \frac{C(t, t)}{t} = \lim_{t \rightarrow 0^+} \frac{\varphi^{[-1]}(2\varphi(t))}{t} = \lim_{x \rightarrow +\infty} \frac{\varphi^{[-1]}(2x)}{\varphi^{[-1]}(x)}$$

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

A continuación mencionamos algunas de las familias de cópulas arquimedias más utilizadas en la literatura. Para cada caso indicaremos la forma de la cópula, su función generatriz, el rango del parámetro, y las dependencias de colas (en ese orden).

■ **Cópula de Clayton**

- $C_\theta(u, v) = [u^{-\theta} + v^{-\theta} - 1]^{-\frac{1}{\theta}}$
- $\varphi_\theta(t) = \frac{1}{\theta} (t^{-\theta} - 1)$
- $\theta \in (0, +\infty)$ .
- $\lambda_U = 0, \lambda_L = 2^{-\frac{1}{\theta}}$ .
- *Observación:* se puede extender el rango del parámetro, para que  $\theta$  pueda pertenecer además a  $[-1, 0)$  con la función generatriz  $\varphi_\theta$ . Sin embargo, bajo esta extensión, la cópula de Clayton pasa a tener la forma

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

deja de ser estricta, y los valores de las dependencias de colas antes mencionados dejan de ser válidos (por ejemplo, si  $\theta = -1$  entonces  $C = W$  y  $\lambda_L = 0$ ).

■ **Cópula de Gumbel**

- $C_\theta(u, v) = \exp \left\{ - \left[ (-\log(u))^\theta + (-\log(v))^\theta \right]^{\frac{1}{\theta}} \right\}$
- $\varphi_\theta(t) = (-\log(t))^\theta$ .
- $\theta \in [1, +\infty)$
- $\lambda_U = 2 - 2^{\frac{1}{\theta}}, \lambda_L = 0$ .

■ **Cópula de Frank**

- $C_\theta(u, v) = -\frac{1}{\theta} \log \left( 1 + \frac{(e^{-\theta u} - 1)(e^{-\theta v} - 1)}{e^{-\theta} - 1} \right)$ .
- $\varphi_\theta(t) = -\log \left( \frac{e^{-\theta t} - 1}{e^{-\theta} - 1} \right)$ .
- $\theta \in \mathbb{R} \setminus \{0\}$ .
- $\lambda_U = 0, \lambda_L = 0$ .

■ **Cópula biparámetro con dependencia de colas independientes**

Esta cópula presenta dependencia de colas inferior y superior no nulas.

- $C_{\alpha, \beta}(u, v) = \left\{ \left[ (u^{-\alpha} - 1)^\beta + (v^{-\alpha} - 1)^\beta \right]^{1/\beta} + 1 \right\}^{-1/\alpha}$
- $\phi_{\alpha, \beta}(t) = (t^{-\alpha} - 1)^\beta$
- $\alpha \in (0, 1], \beta \in [1, +\infty)$ .
- $\lambda_U = 2 - 2^{1/\beta}, \lambda_L = 2^{-1/(\alpha\beta)}$
- $\alpha = -\ln(2 - \lambda_U) / \ln(\lambda_L), \beta = \ln(2) / \ln(2 - \lambda_U)$ .

### 3.2. Estimación de parámetros por máxima verosimilitud.

Esta sección está basada en [4], capítulo 6, secciones 3, 4 y 5.

Supongamos que tenemos cierta cantidad de datos que, sin pérdida de generalidad, asumimos provienen de una cópula. En principio no conocemos nada acerca de la “verdadera” cópula de la cual provienen los datos, así que lo que haremos es buscar familias de cópulas a la cuales podría pertenecer la cópula verdadera, y buscar posibles candidatos dentro de las familias de cópulas que proponemos.

Una vez que tenemos elegida una familia de cópulas a la cual queremos ajustar, debemos encontrar la cópula de la familia que *mejor se ajusta* a los datos. En otras palabras, dentro de todas las cópulas  $C_\theta$  de la familia, queremos encontrar el parámetro  $\hat{\theta}$  que hace que  $C_{\hat{\theta}}$  se ajuste mejor a los datos que el resto de las cópulas de la familia. Supongamos que todas las cópulas de la familia son absolutamente continuas, con densidad  $c_\theta$

Nosotros utilizaremos el método de máxima verosimilitud como ajuste. Supongamos que tenemos una muestra  $\mathbb{X}_n = \{\hat{x}_1, \dots, \hat{x}_n\}$  independiente,

idénticamente distribuída, de tamaño  $n$ , con  $\hat{x}_i \in \mathbf{I}^2$ . Con este método, consideramos un vector aleatorio  $X = \{x_1, \dots, x_n\}$  de tamaño  $n$  que representa el suceso de haber obtenido una muestra; la densidad de ese vector aleatorio (al variar el parámetro  $\theta$ ) es

$$f(x_1, \dots, x_n | \theta) = \prod_{i=1}^n c_\theta(x_i).$$

Desde esta perspectiva, conocemos  $\theta$  y desconocemos  $X$ . Dado que conocemos  $X = \mathbb{X}_n$ , y desconocemos  $\theta$ , la función queda de un parámetro

$$\mathfrak{L}(\theta | \mathbb{X}_n) = \prod_{i=1}^n c_\theta(\hat{x}_i).$$

A la función  $\mathfrak{L}$  se la llama **función de verosimilitud**. Por comodidad, consideraremos el logaritmo de la función de verosimilitud

$$l(\theta | \mathbb{X}_n) = \sum_{i=1}^n \log(c_\theta(\hat{x}_i)).$$

De ahora en más consideramos los siguientes supuestos:

1. Distintos  $\theta$  representan distintas densidades  $c_\theta(x)$ .
2. Todas las densidades tienen el mismo soporte.
3.  $\mathbb{X}_n$  es una muestra aleatoria simple proveniente de  $c_{\theta^0} \ll \lambda$ .
4. Si  $\Omega$  es el espacio de los parámetros, entonces  $\theta^0$  está en el interior de  $\Omega$ .

En el caso  $\Omega \subset \mathbb{R}^n$ , necesitamos además las siguientes suposiciones:

1.  $\theta^0 = (\theta_1^0, \theta_2^0)$  está en el interior de  $\Omega$ , y para casi todo  $x$  en el soporte de  $c$ , se tiene que existe  $\frac{\partial^2}{\partial \theta_1 \partial \theta_2} c_\theta(x)$ , para casi todo  $\theta \in \Omega$ .
2.  $\mathbb{E}_\theta \left[ \frac{\partial}{\partial \theta_1} \log(c_\theta(x)) \right] = \mathbb{E}_\theta \left[ \frac{\partial}{\partial \theta_2} \log(c_\theta(x)) \right] = 0$
3.  $(I_{ij}) = \left( \mathbb{E}_\theta \left[ \frac{\partial}{\partial \theta_i} \log(c_\theta(x)) \frac{\partial}{\partial \theta_j} \log(c_\theta(x)) \right] \right) = \left( \mathbb{E}_\theta \left[ -\frac{\partial^2}{\partial \theta_i \partial \theta_j} \log(c_\theta(x)) \right] \right)$
4.  $I_{ij}$  es finito, y la matriz  $(I_{ij})$  es definida positiva.
5. Supongamos que existen funciones  $M_{ijk}$  tales que

$$\left| \frac{\partial^3}{\partial \theta_i \partial \theta_j \partial \theta_k} \log(c_\theta(u, v)) \right| \leq M_{ijk}(u, v)$$

donde  $\mathbb{E}_{\theta^0}(M_{ijk}) < +\infty$



Con esto, obtenemos el siguiente teorema ([4], página 463, teorema 5.1)

**Teorema 3.2.1.** *Sea  $\mathbb{X}_n$  una muestra aleatoria simple proveniente de una densidad  $c_{\theta^0}$ , con  $\theta^0 \in \mathbb{R}^d$  para algún  $d$ . Asumamos los supuestos antes mencionados. Entonces, con probabilidad tendiendo a 1 con  $n \rightarrow +\infty$  existen soluciones  $\hat{\theta}_n$  del sistema de ecuaciones*

$$\frac{\partial}{\partial \theta_j} l(\theta | \mathbb{X}_n) = 0$$

tales que

1.  $\hat{\theta}_{jn}$  es consistente para estimar  $\theta_j^0$
2.  $\sqrt{n}(\hat{\theta}_n - \theta^0)$  es asintóticamente normal, con esperanza 0 y matriz de covarianza  $(I_{ij}(\theta^0))^{-1}$ .

Las hipótesis del teorema 3.2.1 pueden ser verificadas (basta ver que se puede derivar dentro de la integral, usando por ejemplo la regla de Leibniz), salvo la hipótesis 5, que es muy complicada de verificar por dificultades computacionales.

### 3.3. Bondad de ajuste, test chi cuadrado

Nos basamos en el capítulo 5 de [5].

Una vez estimado el parámetro con máxima verosimilitud, tenemos una cópula  $C_{\hat{\theta}}$  que es candidata a estimar la verdadera cópula  $C$  de la cual provienen nuestros datos. Nos interesa realizar ahora un test de hipótesis

$$\begin{aligned} H_0 : C_{\hat{\theta}} &= C \\ H_1 : C_{\hat{\theta}} &\neq C \end{aligned}$$

en el cual, rechazar  $H_0$  significará que hay evidencia que el ajuste no es bueno.

Nosotros haremos un test chi cuadrado ( $\chi^2$ ) de bondad de ajuste. En este test, dividiremos el cuadrado  $\mathbf{I}^2$  en una grilla uniforme de  $k$  subcuadrados, y llamamos por  $A_i$  ( $i = 1, \dots, k$ ) a cada uno de los subcuadrados. Nos interesa entonces

$$\mathbf{X} = \sum_{i=1}^k \frac{(O_i - E_i)^2}{E_i}$$

donde  $O_i$  es la cantidad de datos contenidos en  $A_i$  y  $E_i$  es la cantidad de datos esperados en  $A_i$  (según  $C$ ). Entonces, si el vector de parámetros  $\theta$  tiene dimensión  $r$ , tenemos que  $\mathbf{X}$  tiene distribución asintótica  $\chi^2$  con  $k - r - 1$  grados de libertad ([5], sección 5.7).

Este test es muy fácil de implementar, y no importa que no estemos trabajando en dimensión 1. Sin embargo, el test no rechazará ninguna cópula  $C^*$  que cumpla  $V_C(A_i) = V_{C^*}(A_i)$  para todo  $i$ . También es complicada la elección del tamaño de la grilla: queremos que sea lo más fina posible (para que el problema anterior no sea tan grave), pero a la vez deseamos que  $O_i$  sea grande para todo  $i$ .

Una vez calculado  $\mathbf{X}$ , podemos hallar el p-valor (la probabilidad de que la variable aleatoria  $\mathbf{X}$  de más grande que la que obtuvimos) y rechazar si es suficientemente chico (digamos 0,05).

## Capítulo 4

# Aplicación a datos bancarios

Este capítulo está basado en [9] y en [10].

Aplicaremos las herramientas que desarrollamos hasta aquí a los datos que tenemos. Estos datos son observaciones diarias de *Credit Default Swaps* (o CDS) de 5 bancos de interés, entre el 1 de enero de 2008 y el 16 de junio de 2010. Varios de los valores, sin embargo, no están disponibles, así que de los 900 datos que deberíamos tener por banco, tenemos entre 450 y 500 (sobre todo después de considerar solo los días donde hay datos para dos bancos al mismo tiempo).

En la primera sección estudiaremos el modelo económico en el que nos encontramos, la definición e implicancia de los CDS en la probabilidad de default de un banco, y cómo nos pueden ayudar a comprender la relación de dependencia entre los activos de dos bancos. En la segunda y tercera sección presentamos los resultados de las estimaciones no paramétricas y paramétricas respectivamente. En la cuarta sección construiremos las matrices *DiDe* presentadas en [9], ampliando las posibilidades de construcción con las distintas cópulas estimadas.

### 4.1. Datos y modelo.

#### 4.1.1. Fórmula de Black Scholes

Supongamos que tenemos una inversión en un activo  $V$ , cuyo precio es modelado a través de un movimiento browniano geométrico

$$V(t) = V(t_0) \left( \exp\{(\mu - 0,5\sigma^2)t + \sigma^2 W(t)\} \right)$$
$$\ln(V(t_1)) - \ln(V(t_0)) | V(t_0) \sim N((\mu - 0,5\sigma^2)(t_1 - t_0), \sigma^2(t_1 - t_0))$$

donde  $\mu$  y  $\sigma$  son constantes. Esta es la solución de la ecuación diferencial estocástica

$$\begin{aligned}dV(t) &= \mu V(t)dt + \sigma V(t)dW(t) \\d\ln(V(t)) &= (\mu - 0,5\sigma^2)dt + \sigma dW(t)\end{aligned}$$

El término  $\mu$  es la tasa de drift, y  $\sigma$  la volatilidad de los retornos de  $V$ . O sea, asumimos retornos log-normales.

Una **opción de compra europea** es un contrato que vincula dos instituciones: beneficiario y vendedor. Al negociar una opción en tiempo  $t_0 = 0$  el beneficiario tiene derecho a comprar una unidad de stock a un precio prefijado  $K$  llamado precio del ejercicio (strike) en un tiempo  $t_1 = T$ .

Si el precio del activo es menor al strike en tiempo  $T$ ,  $V(T) \leq K$ , el beneficiario no ejerce su derecho. De lo contrario se ejecuta la opción y se compra a  $K$ . Por tanto, el beneficiario gana  $(V(T) - K)^+$ .

Como el beneficiario nunca pierde, el contrato debe tener un precio. Black, Scholes y Merton hallan que el precio de tal contrato es

$$C = V(t_0)\Phi(d_1) - Ke^{-r(T-t_0)}\Phi(d_2)$$

donde  $\Phi$  es la función de distribución normal  $(0, 1)$  y

$$\begin{aligned}d_1 &= \frac{\ln(V(t_0)/K) + (r + 0,5\sigma^2)(T - t_0)}{\sigma\sqrt{T - t_0}}, \\d_2 &= d_1 - \sigma\sqrt{T - t_0},\end{aligned}$$

con  $r$  la tasa de interés del activo sin riesgo.

Además, se tiene que la probabilidad de que se ejecute la opción es  $\Phi(-d_2)$  (en realidad, no se usa Black-Scholes para este hecho, solo que los retornos sean log-normales).

#### 4.1.2. Modelo estructural

Supongamos que tenemos una empresa (en este caso van a ser los bancos que estamos estudiando) cuyo valor es  $V(t_0)$ . La empresa tiene una deuda de valor  $\hat{D}$  que debe ser reembolsada en tiempo  $T$ . Dos opciones

- Si  $V(T) \geq \hat{D}$ , entonces la deuda puede ser reembolsada.
- De lo contrario, los acreedores toman posesión de la empresa sin costo, y los poseedores de acciones se quedan sin nada (limited liability).

Lo que queda luego de que la deuda se reembolsa es (“equity value”)

$$C(T) = \max\{V(T) - \hat{D}D, 0\}$$

que puede ser pensado como el valor de una opción call sobre el valor de la empresa, con strike  $\hat{D}D$ .

De manera similar, el valor de la deuda en tiempo  $T$  es

$$DD(T) = V(T) - \max\{V(T) - \hat{D}D, 0\}$$

$$DD(T) = \hat{D}D - \max\{\hat{D}D - V(T), 0\}.$$

El valor de la deuda (con posible default) a tiempo  $T$  se puede pensar (en la segunda ecuación) como el valor de la deuda sin default más una posición corta en una opción put sobre el valor de la empresa con strike  $\hat{D}D$

El precio del equity value es (usando la parte 2)

$$C(t_0) = V(t_0)\Phi(d_1) - e^{-r(T-t_0)}\hat{D}D\Phi(d_2)$$

y el valor de la deuda con posible default es

$$DD(t_0) = V(t_0) - C(t_0) = V(t_0)\Phi(-d_1) + e^{-r(T-t_0)}\hat{D}D\Phi(d_2)$$

donde  $\Phi$  es la función de distribución normal (0, 1) y

$$d_1 = \frac{\ln(V(t_0)/\hat{D}D) + (r + 0,5\sigma^2)(T - t_0)}{\sigma\sqrt{T - t_0}}$$

$$d_2 = d_1 - \sigma\sqrt{T - t_0}$$

y la probabilidad de que se ejecute la opción put y haya quiebra (PoD) es  $\Phi(-d_2)$ .

### 4.1.3. Credit Default Swaps

Un **credit default swap** es un contrato bilateral donde una parte compra protección contra el default de una entidad de referencia.

Supongamos que  $B_1$  le hace un préstamo en tiempo  $t_0$  a  $B_2$  de  $\hat{D}De^{-r(T-t_0)}$  y  $B_2$  debe reembolsar  $\hat{D}D$  a  $B_1$  en tiempo  $T$ . Si  $B_2$  no puede cumplir con el pago (default),  $B_1$  se queda con los activos de  $B_2$  sin costo.

Llamemos  $V(t)$  al valor de los activos de  $B_2$  a tiempo  $t$ . Del enfoque estructural, sabemos que

$$\begin{aligned} C(t_0) &= V(t_0)\Phi(d_1) - e^{-r(T-t_0)}\hat{D}D\Phi(d_2) \\ DD(t_0) &= V(t_0)\Phi(-d_1) + e^{-r(T-t_0)}\hat{D}D\Phi(d_2). \end{aligned}$$

En la segunda ecuación, sumando y restando  $e^{-r(T-t_0)}\hat{D}D$  obtenemos

$$\begin{aligned} DD(t_0) &= e^{-r(T-t_0)}\hat{D}D - \left[ -V(t_0)\Phi(-d_1) + e^{-r(T-t_0)}\hat{D}D\Phi(-d_2) \right] \\ &= e^{-r(T-t_0)}\hat{D}D \left[ 1 - \left( -\frac{1}{d}\Phi(-d_1) + \Phi(-d_2) \right) \right] \\ &= e^{-r(T-t_0)}\hat{D}D [1 - PoD * LGD] \end{aligned}$$

donde PoD es la probabilidad de default de  $B_2$  en tiempo  $T$ , y LGD (loss given default) es cuanto se pierde si hay default, suponemos que es una constante conocida igual a 0,4.

$$\begin{aligned} d &= \frac{e^{-r(T-t_0)}\hat{D}D}{V(t_0)} \\ PoD &= \Phi(-d_2) \\ LGD &= 1 - \frac{1}{d} \frac{\Phi(-d_1)}{\Phi(-d_2)}. \end{aligned}$$

Supongamos que  $B_1$  desea cubrirse del riesgo de que  $B_2$  entre en default; en cuyo caso estaría perdiendo  $(\hat{D}D - V(T))^+$ . Entonces  $B_1$  vende el riesgo a una aseguradora  $A$ : si  $B_2$  entra en default, la aseguradora le paga a  $B_1$  la diferencia  $(\hat{D}D - V(T))^+$  (o compra los activos  $V(T)$  a precio  $\hat{D}D$ ), de lo contrario no se ejecuta la opción. Esto se puede pensar como una opción put, cuyo precio es

$$P = \Phi(-d_2)\hat{D}De^{-r(T-t_0)} - \Phi(-d_1)V(t_0) = \hat{D}De^{-r(T-t_0)} \left[ -\frac{1}{d}\Phi(-d_1) + \Phi(-d_2) \right]$$

Pensaremos entonces al precio de los CDS como el porcentaje respecto a  $e^{-r(T-t_0)}\hat{D}D$  que debe pagar  $B_2$  a  $A$  para cumplir el contrato, o sea

$$CDS * e^{-r(T-t_0)}\hat{D}D = e^{-r(T-t_0)}\hat{D}D [PoD * LGD] \quad (4.1)$$

$$CDS = [PoD * LGD]. \quad (4.2)$$

La fórmula 4.2 relaciona directamente los precios de los credit default swaps con la probabilidad de default. Por lo tanto, estudiando los CDS estamos estudiando la probabilidad de default.

A diferencia de un seguro, cualquier  $B_e$  puede hacer este contrato con  $A$  respecto al préstamo de  $B_1$  a  $B_2$ . En este caso,  $B_e$  obtendrá ganancias solo si  $B_2$  entra en default.

#### 4.1.4. Nuestros datos

Nuestros datos consisten de observaciones diarias de valores de CDS desde el 1º de enero de 2008 hasta el 16 de junio de 2010 (las fechas serán denotadas por el índice  $j$ ). Estos valores se toman para 5 bancos: Citibank NA, Banco Santander Central Hispano SA, BBVA SA, HSBC Bank USA y Credit Lyonnais.

Lo que a nosotros nos interesa son los eventos de que cada uno de estos bancos entre en default en tiempo  $T$  (asumimos que el  $T$  de los contratos de CDS para todos los bancos son similares). Para ello, nos interesará tener información de la variable aleatoria  $V_i(T)$  para cada uno de los bancos ( $i = 1, \dots, 5$ ). Sin embargo, solo tenemos las observaciones diarias de los  $CDS$ .

	Citibank	Santander	BBVA	HSBC	Credit Lyonnais
16/7/08	123.242	48.99	39.688	255.333	48.64
17/7/08	118.407	49.32	38.368	258.333	48.706
18/7/08	#N/A	#N/A	#N/A	#N/A	#N/A
19/7/08	#N/A	#N/A	#N/A	#N/A	#N/A
20/7/08	#N/A	45.087	34.659	231.5	43.929
21/7/08	105.534	46.219	35.494	251	44.474

Cuadro 4.1: Ejemplo de precios de  $CDS$  para los 5 bancos de interés. Para ver los CDS como proporciones (ver 4.1) hay que dividir entre 10000.

Como hay muchos valores no disponibles ( $\#N/A$ ), cuando selecciono un par de bancos para estudiar su dependencia, consideraré las fechas donde estén disponibles los precios de  $CDS$  para ambos bancos a la vez. Aproximadamente, para cada par de bancos, contaremos con 500 observaciones de este estilo.

En el trabajo de Verónica Rodríguez ([9]), se supone que los vectores aleatorios  $(\log(V_i(T)), \log(V_j(T)))$  son vectores aleatorios normales bivariados

con correlación  $\rho_{ij}$ . Se halla la correlación lineal empírica de  $(-\Phi^{-1}(PoD_i), -\Phi^{-1}(PoD_j))$  (donde los  $PoD_i$  son hallados a partir de los  $CDS_i$  usando 4.2), y se utiliza como estimador de  $\rho_{ij}$ . La idea es que podemos estandarizar el vector aleatorio  $(\log(V_i(T)), \log(V_j(T)))$  bajo una transformación lineal, y la correlación  $\rho_{ij}$  no se verá alterada, y además, el vector  $(-\Phi^{-1}(PoD_i), -\Phi^{-1}(PoD_j))$  puede pensarse como  $(\log(\hat{D}D_i)_S, \log(\hat{D}D_j)_S)$ , los umbrales de default en tiempo  $T$  luego de la estandarización.

Nosotros relajaremos estos supuestos. Supondremos que las funciones de distribución conjuntas  $(\log(V_i(T)), \log(V_j(T)))$  son absolutamente continuas (con esto nos aseguramos marginales con funciones de distribución continuas, y existencia de la densidad de la cópula), y la existencia de funciones estrictamente decrecientes  $\alpha_i$  tales que  $\log(\hat{D}D_i) = \alpha_i(CDS_i)$ . Estimamos entonces la cópula del vector aleatorio  $(\log(\hat{D}D_i), \log(\hat{D}D_j))$  (dicha cópula será  $u + v - 1 + C(1 - u, 1 - v)$ , donde  $C$  es la cópula del vector aleatorio  $(CDS_i, CDS_j)$ ), y suponemos que esa cópula será la misma que la del vector aleatorio  $(\log(V_i(T)), \log(V_j(T)))$ .

La idea entonces es estimar la cópula de  $(CDS_i, CDS_j)$  directamente. Para ello aplicamos a cada coordenada su distribución empírica, para así obtener una muestra aleatoria simple con valores en  $\mathbf{I}^2$  (llamaremos de ahora en más a esto “datos estandarizados”).

A la hora de ver las figuras a continuación, hemos de tener en cuenta que los valores altos de CDS corresponden a probabilidades de default altas, por lo tanto el sector de las cópulas cercano al  $(1, 1)$  es el que más nos interesa (aquí se observa la dependencia en momentos de default). De la misma manera, nos interesa mucho más  $\lambda_U$  que  $\lambda_L$ : si  $\lambda_U \neq 0$  entonces es un indicio de dependencia fuerte en momentos de default. Sin embargo,  $\lambda_U = 0$  no necesariamente es un indicador de baja dependencia.

A continuación veremos dos ejemplos, uno donde los datos presentan cierta simetría, y otro donde eso no suceda. Dado que en nuestras estimaciones paramétricas usamos siempre familias que están dentro de las cópulas arquimedianas, la simetría es un detalle importante: no importa qué cópula arquimediana usemos, si los datos no presentan cierta simetría con respecto a la diagonal principal, el ajuste necesariamente será malo.

En el apéndice se mostrarán todos los gráficos y tablas.



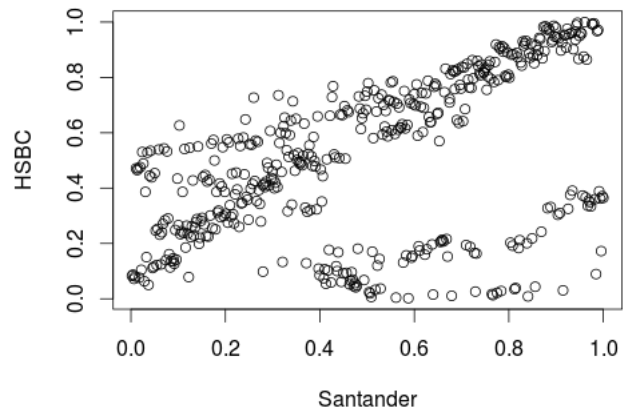


Figura 4.1: Plot de datos estandarizados, Santander - HSBC. Tau de Kendall  $t = 0,39$ . Rho de Spearman  $r = 0,484$ .

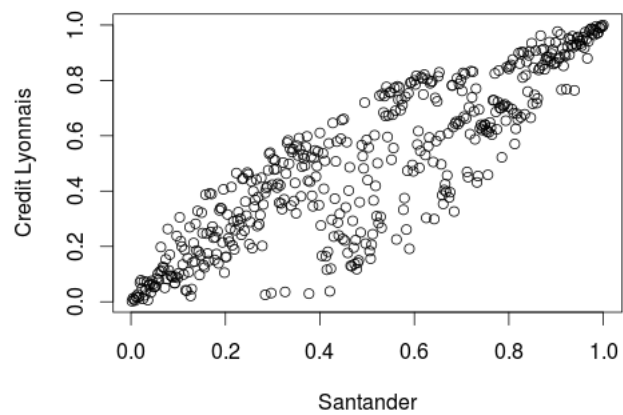


Figura 4.2: Plot de datos estandarizados, Santander - Credit Lyonnais. Tau de Kendall  $t = 0,698$ . Rho de Spearman  $r = 0,854$ .

## 4.2. Resultados: caso no paramétrico

En esta sección presentamos el resultado de aplicar el estimador no paramétrico de random forest a los datos, y algunas observaciones que pueden realizarse a partir de las gráficas.

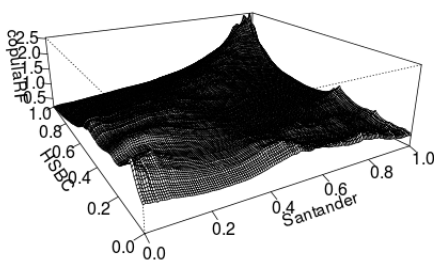


Figura 4.3: Estimación no paramétrica, Santander - HSBC.

En la figura 4.3 podemos observar un máximo en la densidad cerca del  $(1,1)$ , de lo cual deducimos que hay dependencia (concordancia positiva) cuando ambos bancos tienen bajos activos, lo cual indica también que es posible que la cópula presente dependencia de colas superior  $\lambda_U \neq 0$ . De la misma manera observamos que no parece haber máximos cerca de  $(0,0)$ , lo cual es indicio que  $\lambda_L$  podría ser 0. Fuera de los extremos no parece haber mucha regularidad en la dependencia. No se puede esperar un buen ajuste con cópulas arquimedianas, pero dentro de las opciones, las mejores familias para ajustar parecen ser la biparamétrica, o la Gumbel (caso particular de la biparamétrica con  $\alpha \rightarrow 0$ , [15] página 31).

En la figura 4.4 observamos máximos locales cerca de  $(0,0)$  y de  $(1,1)$ , lo cual es indicio de dependencia (concordancia positiva) en los casos donde los bancos tienen altos o bajos activos respectivamente. Estos máximos dan lugar a pensar que puede haber dependencia de colas (superior e inferior) distintas de cero. Las mejores familias para ajustar parecen ser la biparamétrica (en especial cuando haya indicios de dependencia de colas distintas de cero) o la Frank (dependencia de colas ambas iguales a cero).

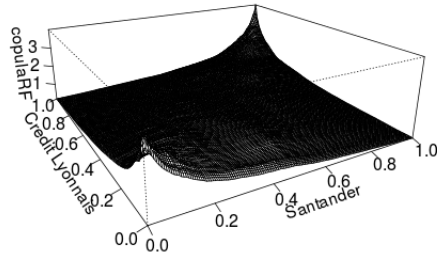


Figura 4.4: Estimación no paramétrica, Santander - Credit Lyonnais

### 4.3. Resultados: caso paramétrico

Realizamos ajuste de parámetros y tests de bondad de ajuste para las familias Gumbel, Frank, Clayton y biparamétrica para cada par de bancos. Para las primeras tres familias, utilizamos el paquete “copula” del software **R**. Para la familia biparamétrica implementamos un código de ajuste de parámetros por máxima verosimilitud, y un test chi cuadrado de bondad de ajuste. Para cada una de las cópulas estimadas, calcularemos las dependencias de colas correspondientes al parámetro obtenido.

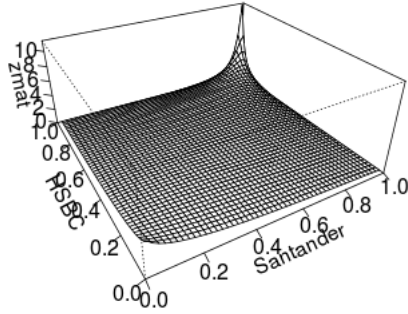


Figura 4.5: Estimación paramétrica, familia Gumbel ( $\theta = 1,55$ ), Santander - HSBC.  $\lambda_U = 0,43$ ,  $\lambda_L = 0$ . P-valor=0.

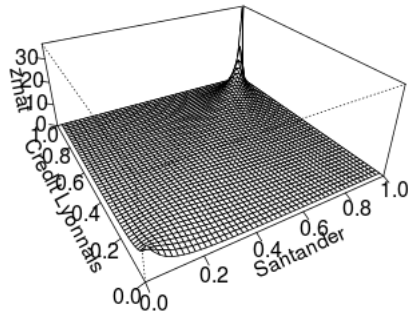


Figura 4.6: Estimación paramétrica, familia Gumbel ( $\theta = 3,31$ ), Santander - Credit Lyonnais.  $\lambda_U = 0,77$ ,  $\lambda_L = 0$ . P-valor= $1,0 \times 10^{-12}$ .

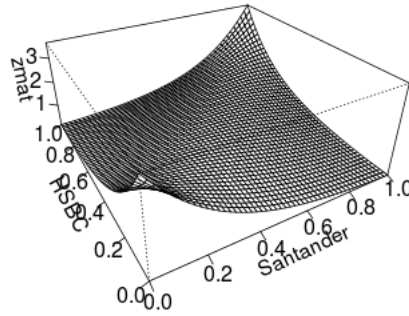


Figura 4.7: Estimación paramétrica, familia Frank ( $\theta = 3,51$ ), Santander - HSBC.  $\lambda_U = 0$ ,  $\lambda_L = 0$ . P-valor=0.

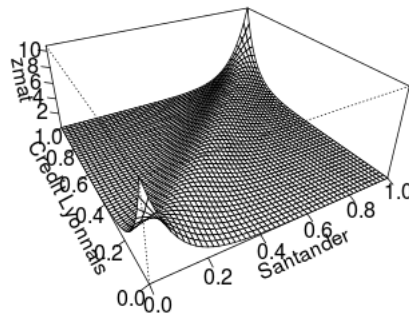


Figura 4.8: Estimación paramétrica, familia Frank ( $\theta = 10,63$ ), Santander - Credit Lyonnais.  $\lambda_U = 0$ ,  $\lambda_L = 0$ . P-valor=0.

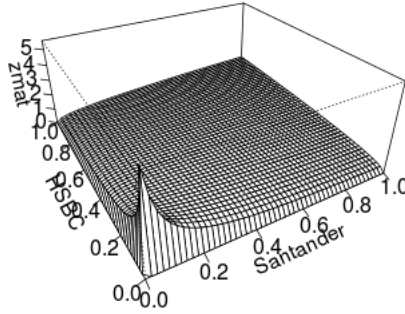


Figura 4.9: Estimación paramétrica, familia Clayton ( $\theta = 0,44$ ) Santander - HSBC.  $\lambda_U = 0$ ,  $\lambda_L = 0,21$ . P-valor=0.

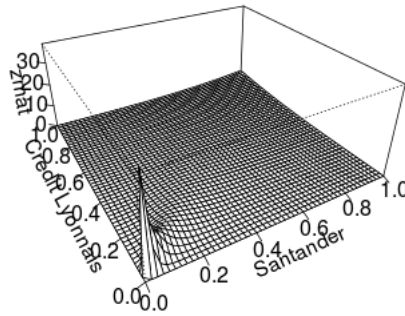


Figura 4.10: Estimación paramétrica, familia Clayton ( $\theta = 2,90$ ) Santander - Credit Lyonnais.  $\lambda_U = 0$ ,  $\lambda_L = 0,79$ . P-valor=0.

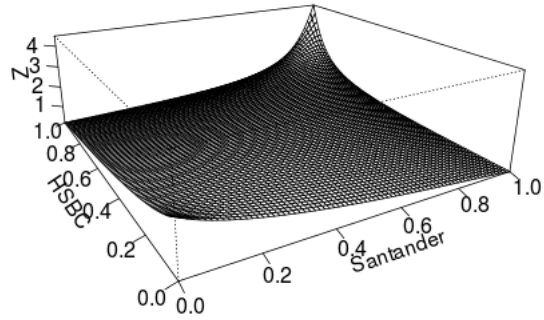


Figura 4.11: Estimación paramétrica, familia Biparamétrica ( $\theta = (1,1 \times 10^{-6}, 1,5)$ ) Santander - HSBC.  $\lambda_U = 0,41$ ,  $\lambda_L = 0$ . P-valor=0.

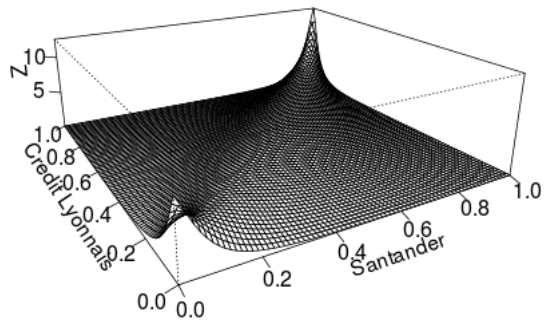


Figura 4.12: Estimación paramétrica, familia Biparamétrica ( $\theta = (0,4, 2,8)$ ) Santander - Credit Lyonnais.  $\lambda_U = 0,72$ ,  $\lambda_L = 0,54$ . P-valor= $1,1 \times 10^{-12}$ .

## 4.4. Indicadores de riesgo bancario

A partir de las cópulas se pueden obtener distintos indicadores de riesgo bancario. Como ejemplo, calcularemos la *Distress Dependence Matrix* (a la cual llamaremos *DiDe*), o sea, la matriz que en la entrada  $(i, j)$  tiene el valor de la probabilidad de que el banco  $i$  entre en default, condicionado al default del banco  $j$ . Para calcular  $DiDe(i, j)$  necesitamos las probabilidades de default de cada banco por separado, y luego la probabilidad conjunta de default.

Si tuviéramos valores de default actuales para cada uno de los bancos, bien se podrían usar esos datos para este cálculo, pero no es nuestro caso. Entonces la idea será fijar un valor de default para cada banco a partir de nuestros datos, bajo cierto criterio. Para cada uno de los bancos tenemos observaciones de los CDS. A partir de ellos, utilizando la fórmula 4.2 podemos obtener las probabilidades de default para cada banco al variar el tiempo. Tomaremos el promedio de estas probabilidades de default para cada banco, y los utilizaremos como valores de referencia. En esta sección, llamaremos a dichos valores  $PoD_i$ .

Para las probabilidades conjuntas de default, idealmente, nos gustaría tener la función de distribución conjunta  $(V_i, V_j)$  (llamémosla  $F(x, y)$ ), donde  $V_i$  es la distribución de los activos del banco  $i$  a tiempo  $T$ . Si conocemos  $F$ , tenemos que la probabilidad conjunta de default es

$$F(\hat{D}D_i, \hat{D}D_j).$$

Sin embargo, si conocemos la cópula subyacente  $C$  de  $F$ , se tiene que

$$F(\hat{D}D_i, \hat{D}D_j) = C(PoD_i, PoD_j),$$

y tenemos varias maneras de estimar  $C$ . Por lo tanto, haremos la matriz *DiDe* para cada una de nuestras maneras de estimar la cópula  $C$  que hemos visto.

*Observación 4.4.1.* Con nuestros cálculos, lo correcto es decir  $F(\hat{D}D_i, \hat{D}D_j) = 1 - PoD_i - PoD_j + C(1 - PoD_i, 1 - PoD_j)$ , ya que valores cercanos a  $(1, 1)$  en  $C$  corresponden a comportamientos cerca de  $-\infty$  en  $F$ . Esto siempre es solucionable estimando las cópulas invirtiendo el orden de los datos luego de estandarizarlos (o sea, si los datos que usamos (luego de estandarizados) los llamamos *data*, hay que estimar la cópula usando  $1 - data$ ).



	PoD
Citibank	0.07
Santander	0.02
BBVA	0.02
HSBC	0.10
Credit Lyonnais	0.02

Cuadro 4.2: Valores de default de referencia

Los valores de default de referencia los tomamos para cada banco por separado.

Dado que no se obtienen buenos valores en los test de ajuste, presentamos aquí a modo de ejemplo la matriz  $DiDe$  para la estimación de la densidad de la cópula por random forests. En todos los casos hay indicios de dependencia de cola superior distinta a cero, por lo cual los valores de la matriz  $DiDe$  en la estimación por random forests pueden ser menores a los verdaderos (ver 2.3.4).

	Citibank	Santander	BBVA	HSBC	Credit Lyonnais
Citibank	1.00	0.17	0.12	0.21	0.15
Santander	0.05	1.00	0.14	0.06	0.15
BBVA	0.04	0.15	1.00	0.04	0.14
HSBC	0.30	0.25	0.19	1.00	0.21
Credit Lyonnais	0.04	0.13	0.12	0.04	1.00

Cuadro 4.3: Matriz  $DiDe$  para estimación no paramétrica por random forests.

# Conclusiones

A continuación enumeramos algunas conclusiones que se desprenden de nuestro trabajo acerca de la teoría de *cóputas*:

- Introdujimos y desarrollamos la teoría de las cóputas. Las cóputas nos permiten comprender la dependencia existente entre dos variables aleatorias dadas, y bajo el supuesto de marginales con funciones de distribución continuas, establecen una biyección entre las funciones de distribución conjuntas  $F_{XY}$  y las ternas  $(F_X, F_Y, C)$  donde  $F_X$  y  $F_Y$  son las marginales de  $F_{XY}$  y  $C$  es la cóputa subyacente (teorema de Sklar).
- Estudiamos como se pueden desprender medidas de dependencia clásicas (tau de Kendall y rho de Spearman) a partir de las cóputas.
- Dada una muestra aleatoria simple  $\mathbb{X}_n \subset \mathbb{R}^2$ , vimos como estimar de manera paramétrica y no paramétrica la cóputa subyacente. En particular, presentamos un estimador no paramétrico de densidades de cóputas basado en *random forests*, que parece ser como mínimo competente frente a la estimación por kernels gaussianos (según las simulaciones realizadas en el software **R**). También introdujimos una familia arquimediana biparamétrica que no tiene restricciones a la hora de medir dependencia de colas (existe una biyección entre el vector de parámetros de la familia y el vector de dependencia de colas superior e inferior).

También nombraremos algunas de las conclusiones que pueden desprenderse del estudio de los datos (en el apéndice están todas las gráficas y tablas).

- Siguiendo [9], estimamos las cóputas de los Credit Default Swaps, y sabemos que esas cóputas representan la estructura de dependencia entre los bancos.

- Las medidas de concordancia obtenidas en todos los datos (tau de Kendall y rho de Spearman) son mayores a cero. En algunos casos se obtienen valores particularmente altos de concordancia. Esto indica que los movimientos bancarios están acompasados: cuando un banco tiene mal desempeño, el resto de los bancos probablemente también lo tenga; de la misma manera, altos rendimientos de un banco harán más probable altos rendimientos en los demás.
- La gran mayoría de los plots de datos (estandarizados) muestran que no hay simetría respecto a la diagonal principal. En dichos casos, debemos esperar malos ajustes de las familias arquimedianas. De hecho, cuando realizamos las estimaciones paramétricas de las cópulas, los p-valores obtenidos en casi todos los casos son casi nulos, exceptuando el caso donde los datos son más simétricos (Santander - BBVA). Dentro de esto, las familias que presentan mejores ajustes son la familia Gumbel y la familia biparamétrica.
- Las estimaciones paramétricas utilizando las familias Gumbel y biparamétrica (las que pueden presentar dependencia de colas superior distinta a cero) dan en todos los casos valores de dependencia de cola superior mayor a cero. Esto es un indicio de fuerte dependencia cuando los bancos están en zona de default.
- Las estimaciones no paramétricas proveen muy buenos resultados en términos de estimar la forma de las cópulas verdaderas: son útiles para ver en qué lugares se dan máximos o mínimos locales o absolutos (permite tener buenas conjeturas para futuras estimaciones paramétricas). Sin embargo, al suavizar la cópula verdadera, no es útil para estimar dependencia de colas.
- En las matrices  $DiDe$ , al no haber buenos ajustes en los modelos (salvo Santander - BBVA), uno tendería a fiarse más de la matriz  $DiDe$  proveniente de la estimación no paramétrica (basándose en el teorema 2.3.1. Sin embargo, dicha matriz tiende a subestimar las probabilidades condicionales (ver la observación 2.3.4).

# Apéndice

En este capítulo presentaremos los resultados que obtuvimos al aplicar los métodos estudiados a los datos que teníamos de Credit Default Swaps. Además adjuntamos los programas en el software **R** que utilizamos. Para algunas tablas nos es más cómodo referenciar a los 5 bancos con números:

1. Citibank NA.
2. Banco Santander Central Hispano SA.
3. BBVA SA.
4. HSBC Bank USA.
5. Credit Lyonnais.

## Plots de datos.

Presentamos a continuación las gráficas de los datos, emparejados dos a dos, y estandarizados aplicando la distribución empírica.

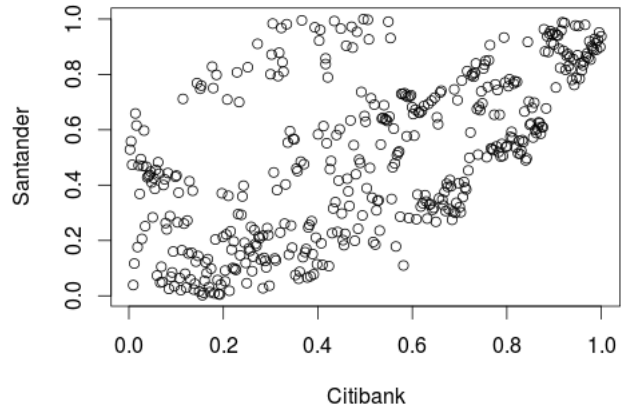


Figura 4.13: Plot de datos estandarizados, Citibank - Santander

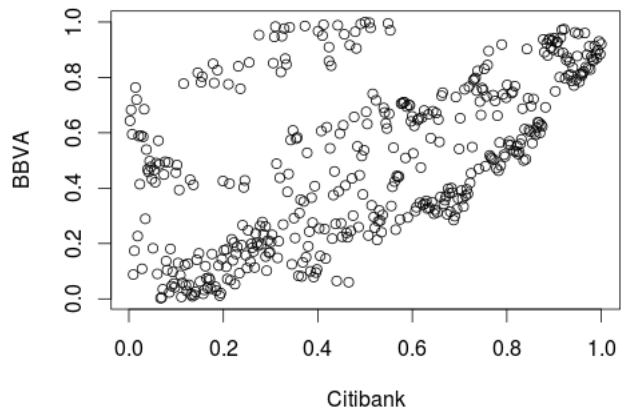
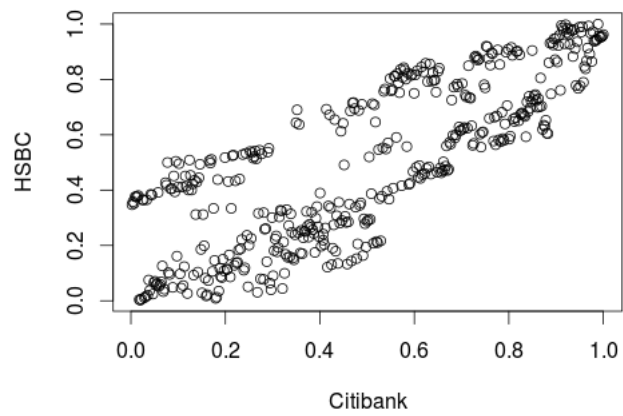


Figura 4.14: Plot de datos estandarizados, Citibank - BBVA



captionPlot de datos estandarizados, Citibank - HSBC

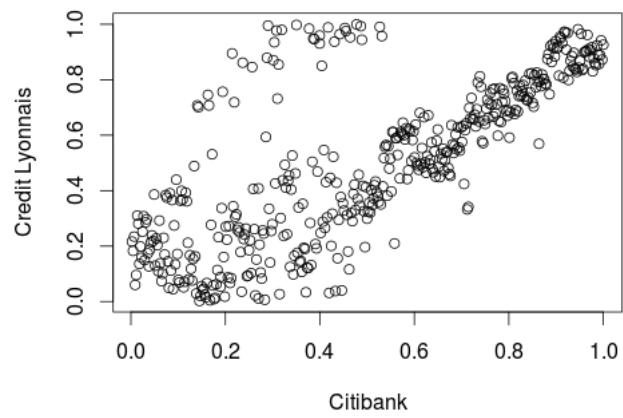


Figura 4.15: Plot de datos estandarizados, Citibank - Credit Lyonnais

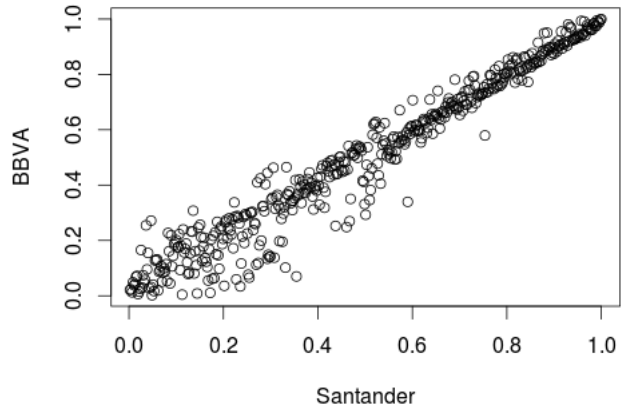


Figura 4.16: Plot de datos estandarizados, Santander - BBVA

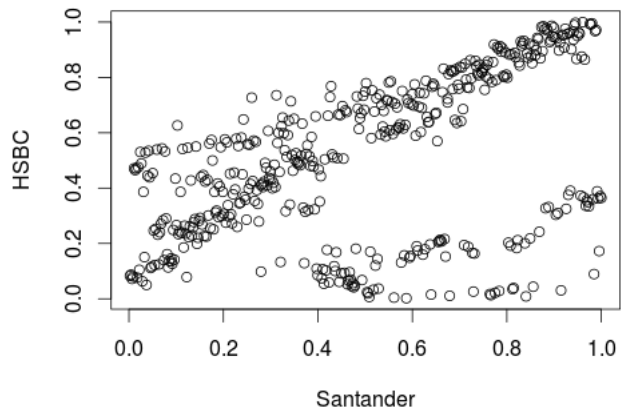


Figura 4.17: Plot de datos estandarizados, Santander - HSBC

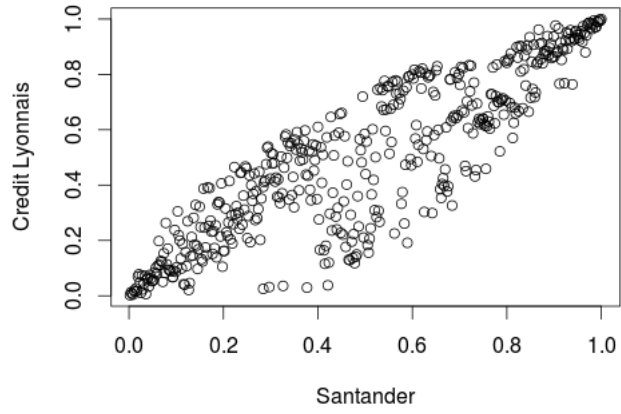


Figura 4.18: Plot de datos estandarizados, Santander - Credit Lyonnais

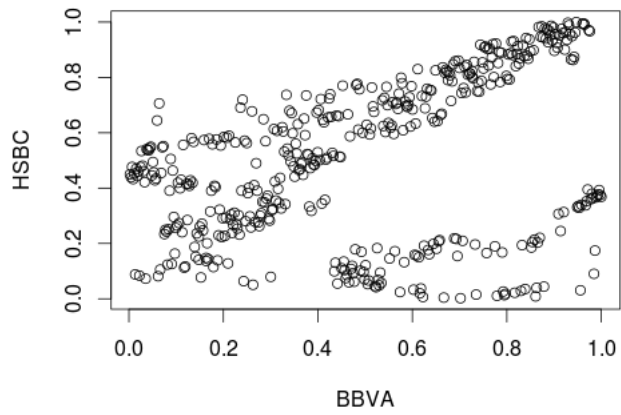


Figura 4.19: Plot de datos estandarizados, BBVA - HSBC



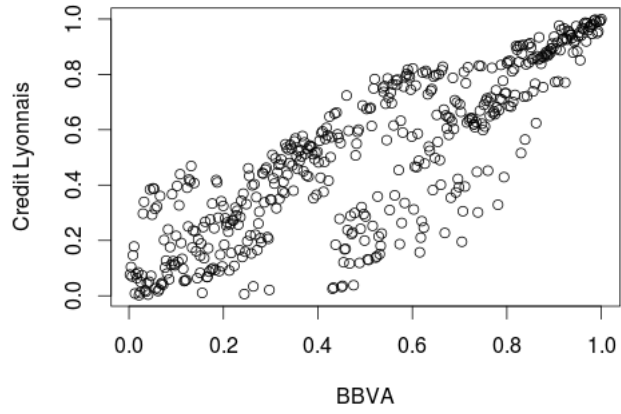


Figura 4.20: Plot de datos estandarizados, BBVA - Credit Lyonnais

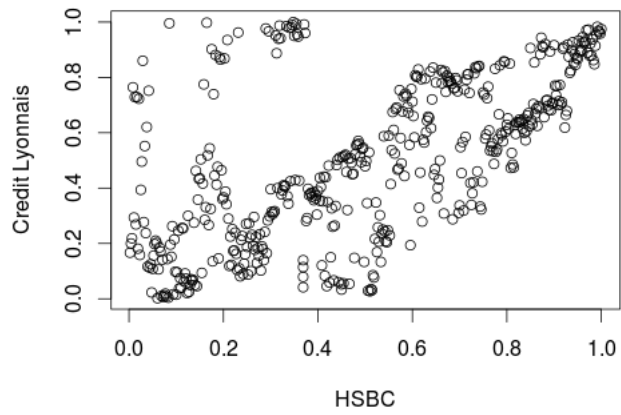


Figura 4.21: Plot de datos estandarizados, HSBC - Credit Lyonnais

## Estimación de concordancia

La siguiente tabla tiene los valores estimados de la tau de Kendall y de la rho de Spearman, utilizando la cópula empírica (2.1.3).

	1-2	1-3	1-4	1-5	2-3	2-4	2-5	3-4	3-5	4-5
Tau Kendall	0.44	0.43	0.60	0.58	0.86	0.39	0.70	0.31	0.65	0.46
Rho Spearman	0.61	0.57	0.81	0.74	0.95	0.48	0.85	0.39	0.81	0.61

Cuadro 4.4: Estimación de concordancias. Valores estimados de la tau de Kendall y la rho de Spearman utilizando la cópula empírica.

## Matrices *DiDe*

Presentamos a continuación los valores de default de referencia para cada banco (ver sección 4.4) y luego las distintas matrices *DiDe* para las distintas estimaciones de cópulas que realizamos (no paramétrica por random forests y las cuatro familias paramétricas arquimedianas).

	PoD
Citibank	0.07
Santander	0.02
BBVA	0.02
HSBC	0.10
Credit Lyonnais	0.02

Cuadro 4.5: Valores de default de referencia

	Citibank	Santander	BBVA	HSBC	Credit Lyonnais
Citibank	1.00	0.17	0.12	0.21	0.15
Santander	0.05	1.00	0.14	0.06	0.15
BBVA	0.04	0.15	1.00	0.04	0.14
HSBC	0.30	0.25	0.19	1.00	0.21
Credit Lyonnais	0.04	0.13	0.12	0.04	1.00

Cuadro 4.6: Matriz *DiDe* para estimación no paramétrica por random forests.

	Citibank	Santander	BBVA	HSBC	Credit Lyonnais
Citibank	1.00	0.69	0.68	0.39	0.71
Santander	0.21	1.00	0.43	0.16	0.46
BBVA	0.22	0.45	1.00	0.17	0.47
HSBC	0.55	0.75	0.74	1.00	0.76
Credit Lyonnais	0.19	0.41	0.40	0.15	1.00

Cuadro 4.7: Matriz *DiDe* para las estimaciones biparamétricas.

	Citibank	Santander	BBVA	HSBC	Credit Lyonnais
Citibank	1.00	0.75	0.74	0.42	0.77
Santander	0.23	1.00	0.48	0.18	0.52
BBVA	0.24	0.50	1.00	0.18	0.53
HSBC	0.60	0.80	0.80	1.00	0.82
Credit Lyonnais	0.21	0.45	0.44	0.16	1.00

Cuadro 4.8: Matriz  $DiDe$  para las estimaciones Gumbel.

	Citibank	Santander	BBVA	HSBC	Credit Lyonnais
Citibank	1.00	0.28	0.28	0.25	0.29
Santander	0.09	1.00	0.10	0.08	0.10
BBVA	0.09	0.10	1.00	0.09	0.10
HSBC	0.35	0.38	0.38	1.00	0.38
Credit Lyonnais	0.08	0.09	0.09	0.07	1.00

Cuadro 4.9: Matriz  $DiDe$  para las estimaciones Frank.

	Citibank	Santander	BBVA	HSBC	Credit Lyonnais
Citibank	1.00	0.12	0.12	0.12	0.12
Santander	0.04	1.00	0.04	0.04	0.04
BBVA	0.04	0.04	1.00	0.04	0.04
HSBC	0.17	0.17	0.17	1.00	0.17
Credit Lyonnais	0.03	0.03	0.03	0.03	1.00

Cuadro 4.10: Matriz  $DiDe$  para las estimaciones Clayton.

## Estimaciones no paramétricas por random forests

A continuación presentamos las estimaciones de las densidades de cópulas por random forests. Las simulaciones se corrieron con grillado de  $150 \times 150$ , 50 particiones (aleatorias) de  $\mathbf{I}^2$ , y 2000 árboles.

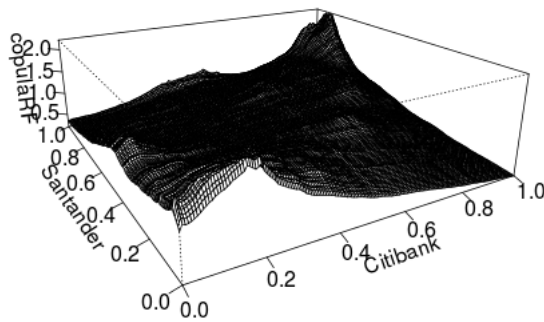


Figura 4.22: Estimación de la cópula de los datos Citibank - Santander utilizando random forests

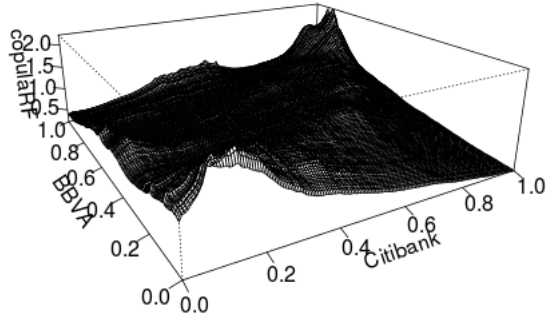


Figura 4.23: Estimación de la cópula de los datos Citibank - BBVA utilizando random forests

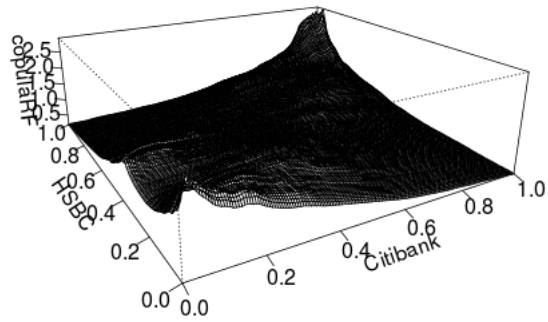


Figura 4.24: Estimación de la cópula de los datos Citibank - HSBC utilizando random forests

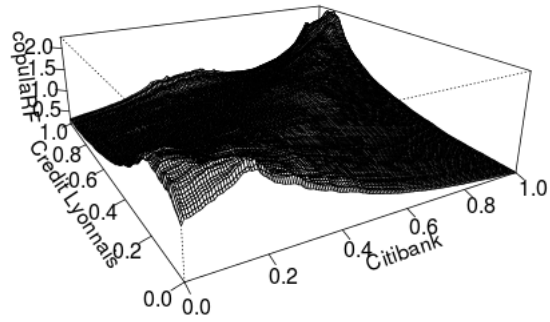


Figura 4.25: Estimación de la cópula de los datos Citibank - Credit Lyonnais utilizando random forests

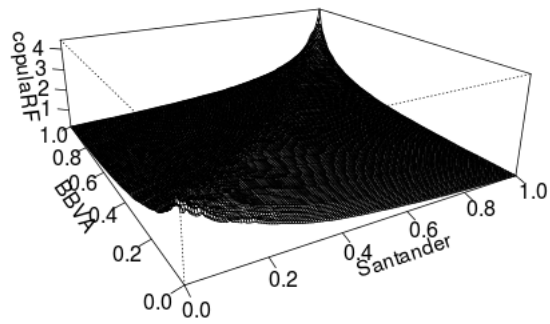


Figura 4.26: Estimación de la cópula de los datos Santander - BBVA utilizando random forests

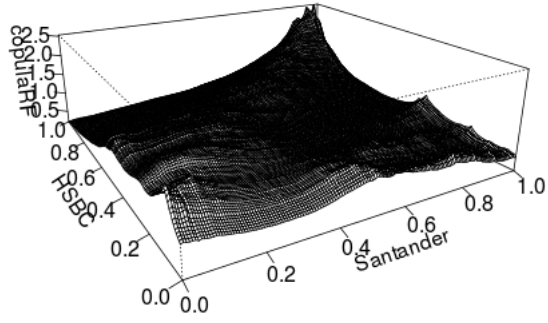


Figura 4.27: Estimación de la cópula de los datos Santander - HSBC utilizando random forests

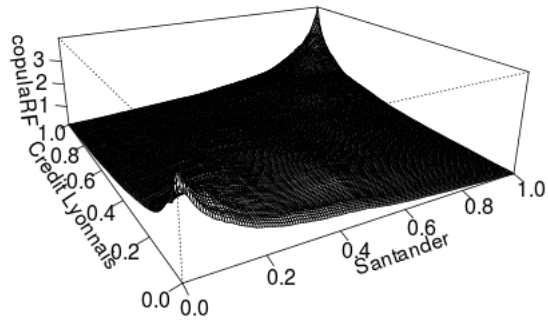


Figura 4.28: Estimación de la cópula de los datos Santander - Credit Lyonnais utilizando random forests



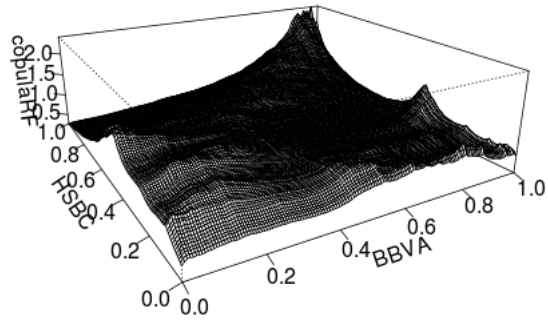


Figura 4.29: Estimación de la cópula de los datos BBVA - HSBC utilizando random forests

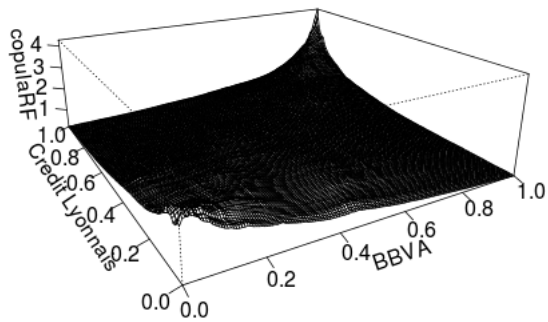


Figura 4.30: Estimación de la cópula de los datos BBVA - Credit Lyonnais utilizando random forests

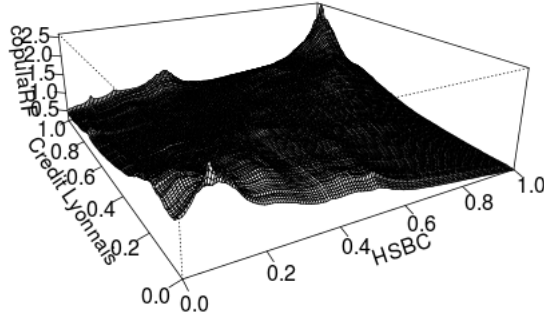


Figura 4.31: Estimación de la cópula de los datos HSBC - Credit Lyonnais utilizando random forests

## Test de bondad de ajuste para familias paramétricas

La tabla a continuación muestra los resultados del test chi cuadrado de bondad de ajuste. Estos test se han realizado dividiendo el cuadrado  $\mathbf{I}^2$  en una grilla de  $10 \times 10$ .

	1-2	1-3	1-4	1-5	2-3	2-4	2-5	3-4	3-5	4-5
Biparamétrica	0.00	0.00	0.00	0.00	0.78	0.00	0.00	0.00	0.00	0.00
Gumbel	0.00	0.00	0.00	0.00	0.83	0.00	0.00	0.00	0.00	0.00
Frank	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Clayton	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

Cuadro 4.11:  $p$ -valores del test chi cuadrado de bondad de ajuste de familias paramétricas.

Obtenemos buenos ajustes únicamente para el par Santander - BBVA, utilizando las familias biparamétrica y Gumbel.

## Estimación por familia biparamétrica

Presentamos los resultados de la estimación paramétrica de las cópulas utilizando la familia arquimediana biparamétrica. Primero presentamos los parámetros estimados, luego las dependencias de colas que inducen dichos parámetros, y luego las gráficas de las densidades.

	1-2	1-3	1-4	1-5	2-3	2-4	2-5	3-4	3-5	4-5
alpha	0.13	0.26	0.13	0.17	0.00	0.00	0.38	0.00	0.00	0.19
beta	1.48	1.35	2.02	1.74	6.17	1.70	2.85	1.35	2.87	1.53

Cuadro 4.12: Parámetros estimados por máxima verosimilitud

	1-2	1-3	1-4	1-5	2-3	2-4	2-5	3-4	3-5	4-5
inferior	0.03	0.14	0.07	0.10	0.00	0.00	0.53	0.00	0.00	0.10
superior	0.40	0.33	0.59	0.51	0.88	0.50	0.72	0.33	0.73	0.43

Cuadro 4.13: Dependencia de colas con los parámetros estimados

En todos los casos obtenemos dependencias de cola superior mayores a cero, lo cual parece indicar (a pesar del mal ajuste en la mayoría de los casos) que hay alta dependencia cuando los bancos se encuentran cerca de default.

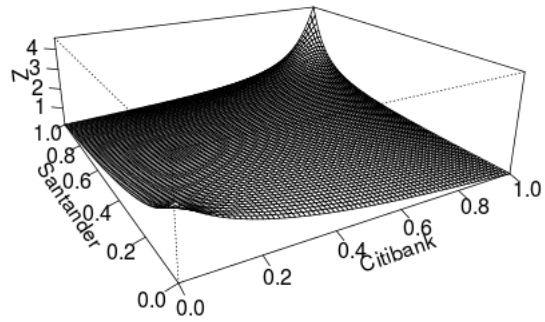


Figura 4.32: Estimación de la cópula de los datos Citibank - Santander utilizando la familia biparamétrica

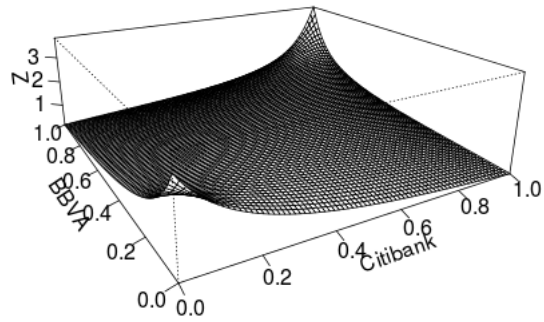


Figura 4.33: Estimación de la cópula de los datos Citibank - BBVA utilizando la familia biparamétrica

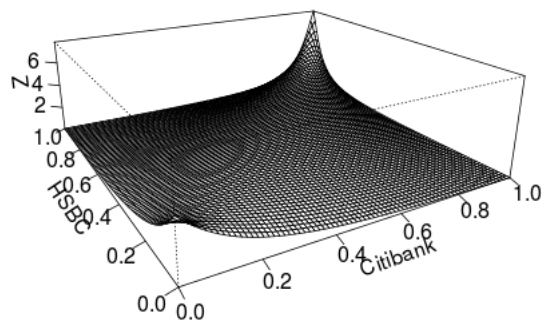


Figura 4.34: Estimación de la cópula de los datos Citibank - HSBC utilizando la familia biparamétrica

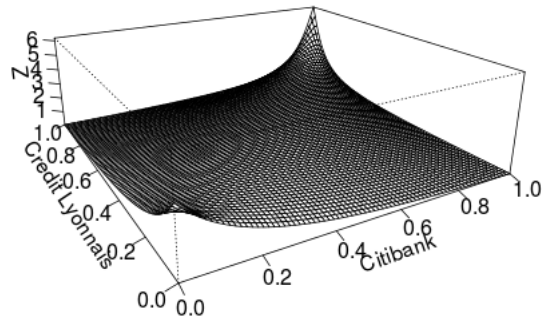


Figura 4.35: Estimación de la cópula de los datos Citibank - Credit Lyonnais utilizando la familia biparamétrica

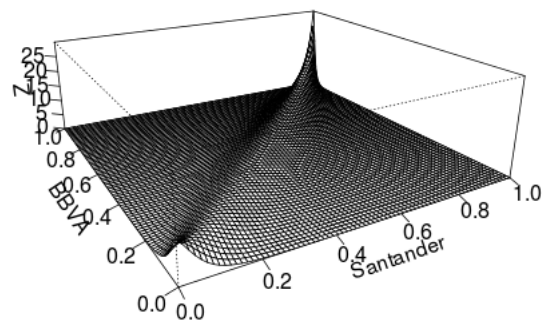


Figura 4.36: Estimación de la cópula de los datos Santander - BBVA utilizando la familia biparamétrica

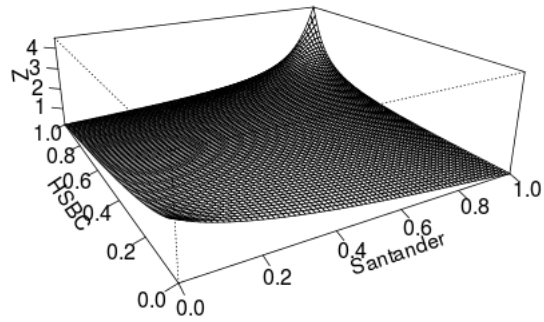


Figura 4.37: Estimación de la cópula de los datos Santander - HSBC utilizando la familia biparamétrica

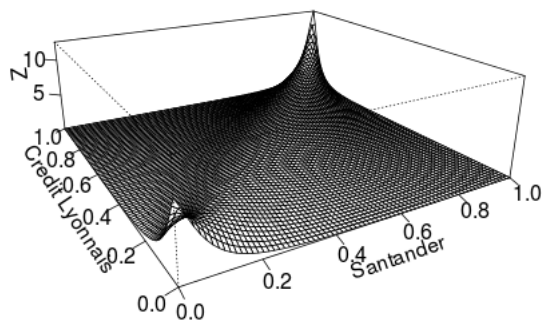


Figura 4.38: Estimación de la cópula de los datos Santander - Credit Lyonnais utilizando la familia biparamétrica

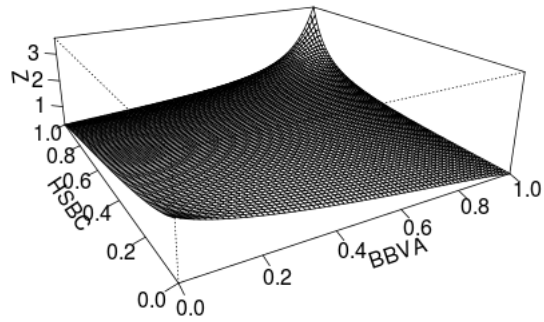


Figura 4.39: Estimación de la cópula de los datos BBVA - HSBC utilizando la familia biparamétrica

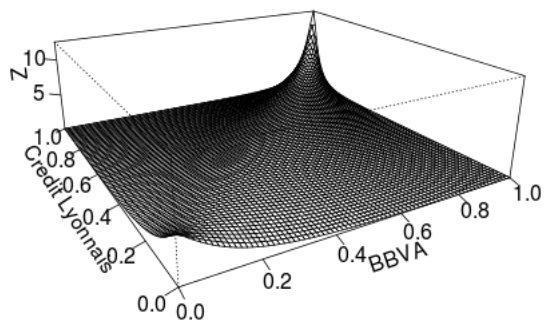


Figura 4.40: Estimación de la cópula de los datos BBVA - Credit Lyonnais utilizando la familia biparamétrica



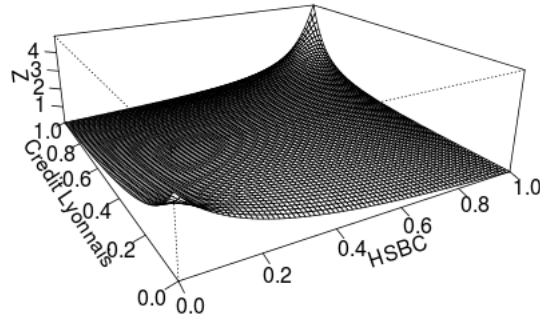


Figura 4.41: Estimación de la cópula de los datos HSBC - Credit Lyonnais utilizando la familia biparamétrica

## Estimación por familia Gumbel

En esta sección presentamos los resultados de las estimaciones paramétricas utilizando la familia arquimediana Gumbel. Presentaremos primero los parámetros estimados, luego los valores de dependencia de colas superior  $\lambda_U$  que inducen dichos parámetros (recordamos  $\lambda_L = 0$  en la familia Gumbel), y por último mostraremos las gráficas de las densidades de las cópulas estimadas.

	1-2	1-3	1-4	1-5	2-3	2-4	2-5	3-4	3-5	4-5
theta	1.56	1.50	2.13	1.87	6.56	1.51	3.31	1.35	2.88	1.65

Cuadro 4.14: Parámetros estimados utilizando la familia Gumbel.

	1-2	1-3	1-4	1-5	2-3	2-4	2-5	3-4	3-5	4-5
lambda	0.44	0.41	0.61	0.55	0.89	0.42	0.77	0.33	0.73	0.48

Cuadro 4.15: Dependencia de colas superior para las cópulas estimadas.

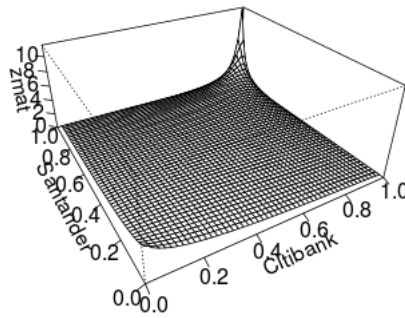


Figura 4.42: Estimación de la cópula de los datos Citibank - Santander utilizando la familia Gumbel

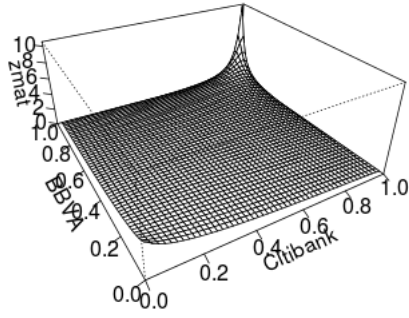


Figura 4.43: Estimación de la cópula de los datos Citibank - BBVA utilizando la familia Gumbel

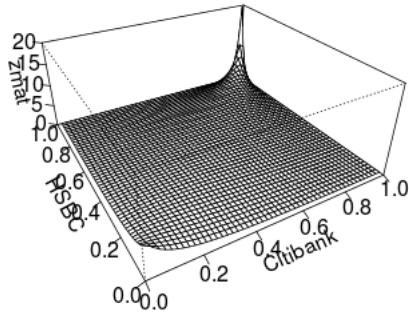


Figura 4.44: Estimación de la cópula de los datos Citibank - HSBC utilizando la familia Gumbel

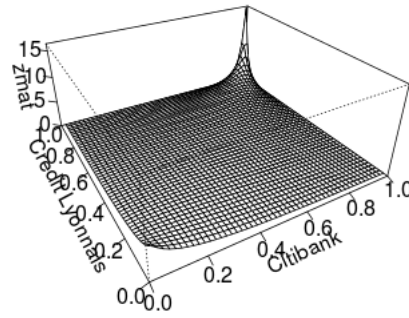


Figura 4.45: Estimación de la cópula de los datos Citibank - Credit Lyonnais utilizando la familia Gumbel

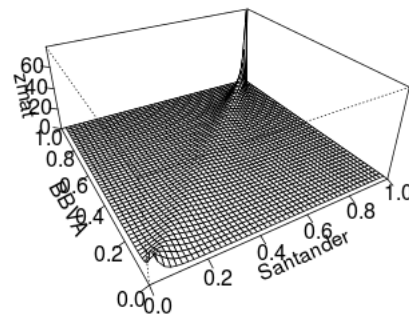


Figura 4.46: Estimación de la cópula de los datos Santander - BBVA utilizando la familia Gumbel

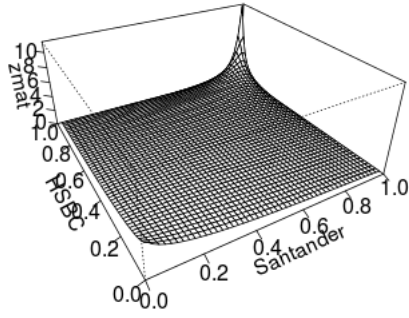


Figura 4.47: Estimación de la cópula de los datos Santander - HSBC utilizando la familia Gumbel

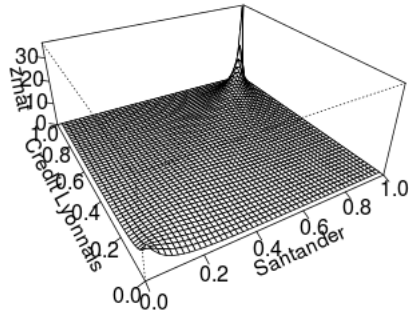


Figura 4.48: Estimación de la cópula de los datos Santander - Credit Lyonnais utilizando la familia Gumbel

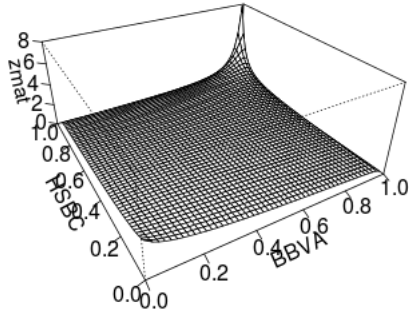


Figura 4.49: Estimación de la cópula de los datos BBVA - HSBC utilizando la familia Gumbel

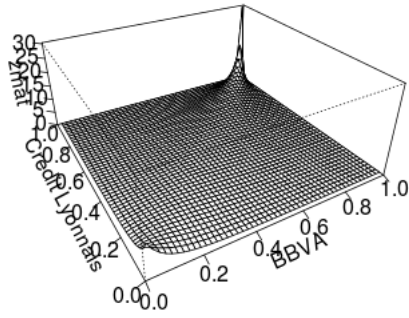


Figura 4.50: Estimación de la cópula de los datos BBVA - Credit Lyonnais utilizando la familia Gumbel

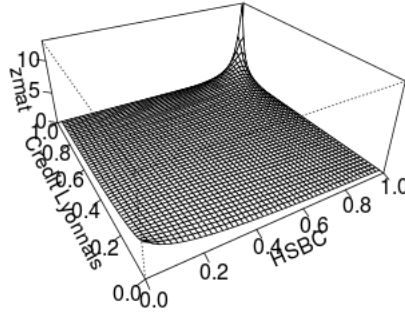


Figura 4.51: Estimación de la cópula de los datos HSBC - Credit Lyonnais utilizando la familia Gumbel

## Estimación por familia Frank

En esta sección presentamos las estimaciones utilizando la familia arquimediana Frank. Recordamos que las cópulas de esta familia tienen dependencia de colas superior e inferior igual a 0. Presentamos los parámetros estimados, y luego las gráficas de las densidades de las cópulas estimadas.

	1-2	1-3	1-4	1-5	2-3	2-4	2-5	3-4	3-5	4-5
theta	4.47	4.19	7.42	7.05	28.85	3.51	10.63	2.59	9.10	5.08

Parámetros estimados utilizando la familia Frank.

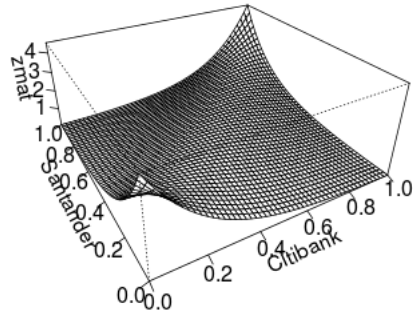


Figura 4.52: Estimación de la cópula de los datos Citibank - Santander utilizando la familia Frank



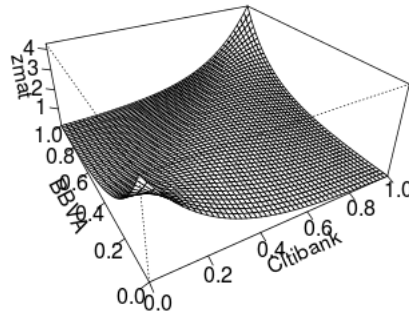


Figura 4.53: Estimación de la cópula de los datos Citibank - BBVA utilizando la familia Frank

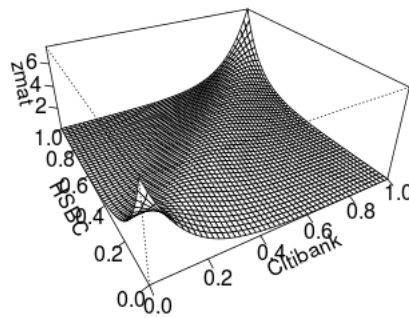


Figura 4.54: Estimación de la cópula de los datos Citibank - HSBC utilizando la familia Frank

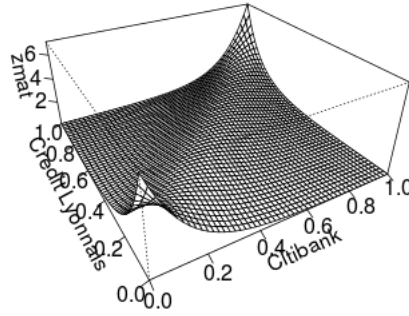


Figura 4.55: Estimación de la cópula de los datos Citibank - Credit Lyonnais utilizando la familia Frank

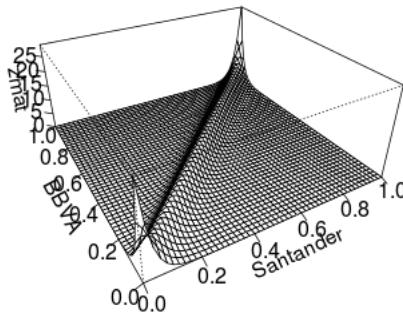


Figura 4.56: Estimación de la cópula de los datos Santander - BBVA utilizando la familia Frank

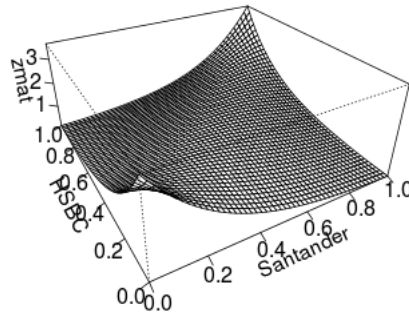


Figura 4.57: Estimación de la cópula de los datos Santander - HSBC utilizando la familia Frank

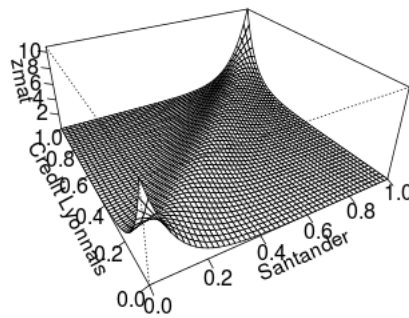


Figura 4.58: Estimación de la cópula de los datos Santander - Credit Lyonnais utilizando la familia Frank

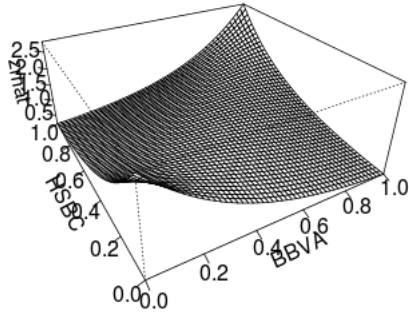


Figura 4.59: Estimación de la cópula de los datos BBVA - HSBC utilizando la familia Frank

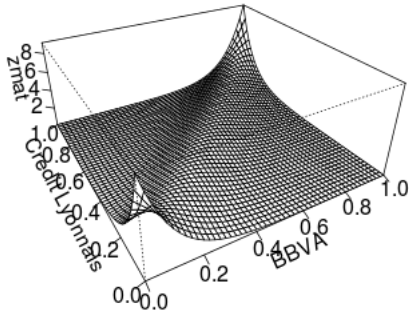


Figura 4.60: Estimación de la cópula de los datos BBVA - Credit Lyonnais utilizando la familia Frank

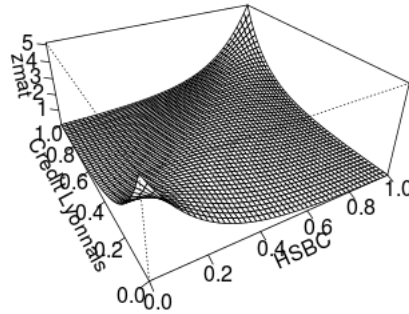


Figura 4.61: Estimación de la cópula de los datos HSBC - Credit Lyonnais utilizando la familia Frank

## Estimación por familia Clayton

Presentamos ahora las estimaciones de las cópulas utilizando la familia arquimediana Clayton. La primera tabla corresponde a la estimación de los parámetros, y la segunda tabla a la dependencia de colas inferior inducida por los parámetros (recordemos que en la familia Clayton,  $\lambda_U = 0$ ).

	1-2	1-3	1-4	1-5	2-3	2-4	2-5	3-4	3-5	4-5
theta	0.75	0.76	1.39	1.05	3.86	0.44	2.90	0.21	1.73	0.86

Cuadro 4.16: Parámetros estimados con la familia Clayton.

	1-2	1-3	1-4	1-5	2-3	2-4	2-5	3-4	3-5	4-5
lambda	0.40	0.40	0.61	0.52	0.84	0.21	0.79	0.03	0.67	0.45

Cuadro 4.17: Dependencia de colas inferior para las cópulas estimadas.

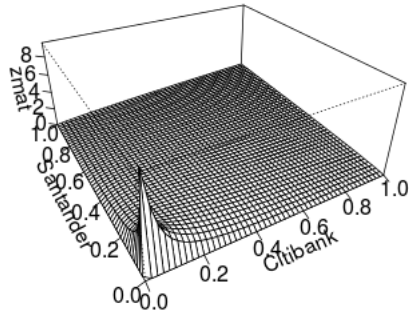


Figura 4.62: Estimación de la cópula de los datos Citibank - Santander utilizando la familia Clayton

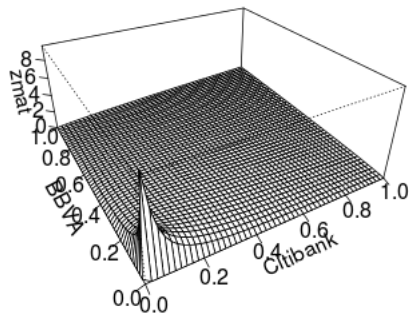


Figura 4.63: Estimación de la cópula de los datos Citibank - BBVA utilizando la familia Clayton

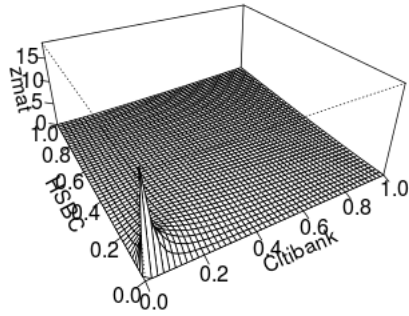


Figura 4.64: Estimación de la cópula de los datos Citibank - HSBC utilizando la familia Clayton

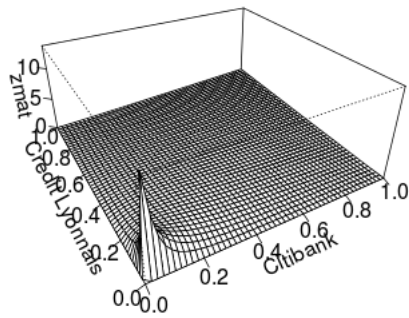


Figura 4.65: Estimación de la cópula de los datos Citibank - Credit Lyonnais utilizando la familia Clayton

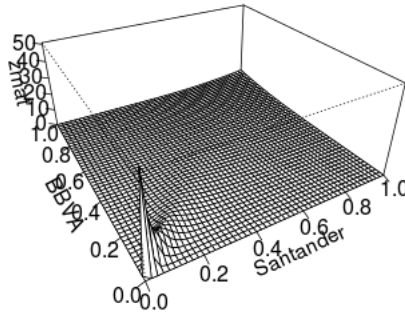


Figura 4.66: Estimación de la cópula de los datos Santander - BBVA utilizando la familia Clayton

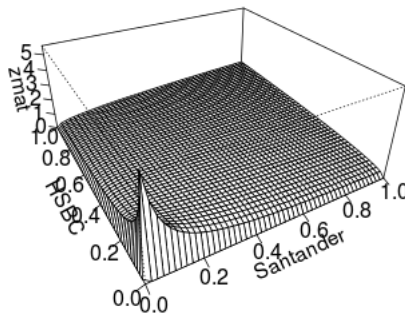


Figura 4.67: Estimación de la cópula de los datos Santander - HSBC utilizando la familia Clayton



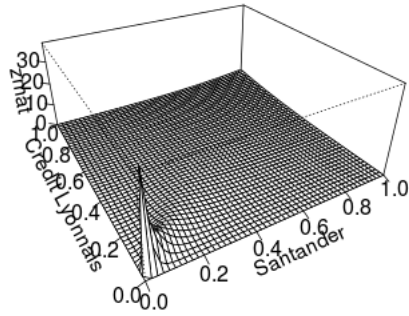


Figura 4.68: Estimación de la cópula de los datos Santander - Credit Lyonnais utilizando la familia Clayton

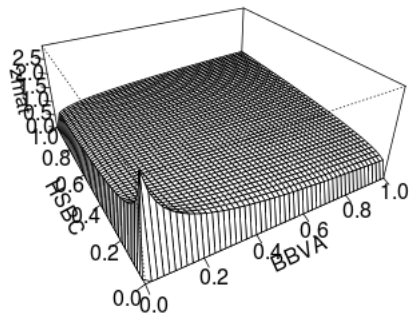


Figura 4.69: Estimación de la cópula de los datos BBVA - HSBC utilizando la familia Clayton

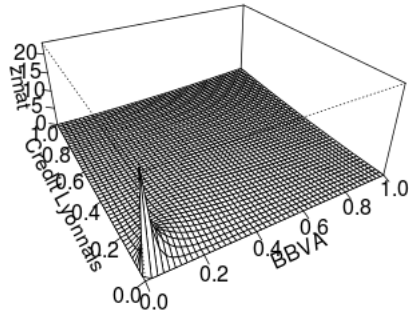


Figura 4.70: Estimación de la cópula de los datos BBVA - Credit Lyonnais utilizando la familia Clayton

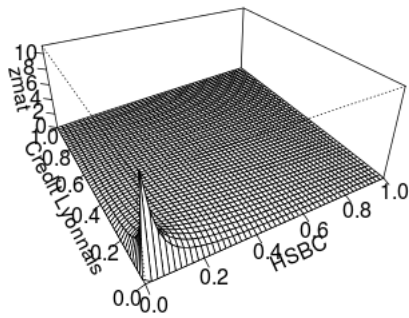


Figura 4.71: Estimación de la cópula de los datos HSBC - Credit Lyonnais utilizando la familia Clayton

# Bibliografía

- [1] R. Nelsen. *An introduction to copulas. Second edition.* Springer Series in Statistics. Springer, New York, 2006. xiv+269 pp. ISBN: 978-0387-28659-4; 0-387-28659-4
- [2] G. Biau, L. Devroye, G. Lugosi. *Consistency of random forests and other averaging classifiers.* J. Mach. Learn. Res. 9 (2008), 2015-2033.
- [3] R. Nelsen. *Dependence and order in families of Archimedean copulas.* J. Multivariate Anal. 60 (1997), no. 1, 111-122.
- [4] E. Lehmann, G. Casella. *Theory of point estimation. Second edition.* Springer Texts in Statistics. Springer-Verlag, New York, 1998. xxvi+589 pp. ISBN: 0-387-98502-6.
- [5] E. Lehmann. *Elements of large-sample theory.* Springer Texts in Statistics. Springer-Verlag, New York, 1999. xii+631 pp. ISBN: 0-387-98595-6
- [6] L. Devroye, L. Györfi, G. Lugosi. *A probabilistic theory of pattern recognition.* Applications of Mathematics (New York), 31. Springer-Verlag, New York, 1996. xvi+636 pp. ISBN: 0-387-94618-7.
- [7] L. Devroye. *A note on the height of binary search trees.* Journal of the ACM, 33:489-498,1986.
- [8] S. X. Chen, T. M. Huang. *Nonparametric estimation for copula functions for dependence modelling.* Canad. J. Statist. 35 (2007), no. 2, 265-282.
- [9] V. Rodriguez. *Indicadores de Riesgo derivados de Depósitos Bancarios Constituidos en el Exterior.* Tesis de Magister en Economía Internacional. Departamento de Economía, Facultad de Ciencias Sociales, Universidad de la República. 10 de marzo de 2010.

- [10] U. Cherubini, E. Luciano, W. Vecchiato. *Copula methods in finance*. Wiley Finance Series. John Wiley and Sons, Ltd., Chichester, 2004. xvi+293 pp. ISBN: 0-470-86344-7.
- [11] G. Frahm, M. Junker, R. Schmidt. *Estimating the tail-dependence coefficient: properties and pitfalls*. Insurance Math. Econom. 37 (2005), no. 1, 80-100.
- [12] D. Li. *On default correlation: a copula function approach*. RiskMetrics Groups. 44 Wall Street. New York, NY 10005. September 16, 1999.
- [13] A. Sklar, *Fonctions de répartition à n dimensions et leurs marges*. Publ Inst Statist Univ Paris (1959), 8:229-231
- [14] W. Hoeffding, *Masstabinvariante Korrelationstheorie*. Schriften des Mathematischen Instituts und des Instituts für Angewandte Mathematik der Universität Berlin 5 Heft (1940), 3:179-233 [Reprinted as Scale-invariant correlation theory. In: Fisher NI, Sen PK (eds) The Collected Works of Wassily Hoeffding. Springer, New York pp 57-107]
- [15] B. Vaz de Melo, R. Martins de Souza, *Measuring Financial risks with copulas*. International Review of Financial Analysis 13 (2004), 27-45.