



UNIVERSIDAD NACIONAL AUTÓNOMA  
DE MÉXICO

---

---

FACULTAD DE CIENCIAS

Cómulas y funciones de dependencia extrema aplicadas al  
cálculo del VaR de datos financieros

T E S I S

QUE PARA OBTENER EL TÍTULO DE:

Actuaría

PRESENTA:

Natalia Sarabia Vásquez

TUTOR

Dr. Yuri Salazar Flores



Ciudad Universitaria, Ciudad de México, 2017.



Universidad Nacional  
Autónoma de México



## **UNAM – Dirección General de Bibliotecas Tesis Digitales Restricciones de uso**

### **DERECHOS RESERVADOS © PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL**

Todo el material contenido en esta tesis está protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor.

**1. Datos del alumno**

Apellido paterno	Sarabia
Apellido materno	Vásquez
Nombre(s)	Natalia
Teléfono	58960839
Universidad Nacional Autónoma de México	Universidad Nacional Autónoma de México
Facultad de Ciencias	Facultad de Ciencias
Carrera	Actuaría
Número de cuenta	30927977-4

**2. Datos del tutor**

Grado	Dr.
Nombre(s)	Yuri
Apellido paterno	Salazar
Apellido materno	Flores

**3. Datos del sinodal 1**

Grado	Dra.
Nombre(s)	Ana
Apellido paterno	Meda
Apellido materno	Guardiola

**4. Datos del sinodal 2**

Grado	Dr.
Nombre(s)	Adán
Apellido paterno	Díaz
Apellido materno	Hernández

**5. Datos del sinodal 3**

Grado	M. en C.
Nombre(s)	Laura Nayeli
Apellido paterno	Liljehult
Apellido materno	León

**6. Datos del sinodal 4**

Grado	M. en C.
Nombre(s)	Daniel
Apellido paterno	Cervantes
Apellido materno	Filoteo

**7. Datos del trabajo escrito**

Título	Cópulas y funciones de dependencia extrema
Subtítulo	aplicadas al cálculo del VaR de datos financieros
Número de páginas	93 págs.
Año	2017

# Índice general

1.	Introducción . . . . .	2
2.	Preliminares . . . . .	3
2.1.	Cóputas . . . . .	8
2.2.	Teorema de Sklar . . . . .	9
2.3.	Densidades de cóputas . . . . .	14
2.4.	Cotas para cóputas . . . . .	14
2.5.	Cóputas fundamentales . . . . .	16
2.6.	Cóputas de supervivencia . . . . .	19
2.7.	Simetría . . . . .	21
2.8.	Conclusiones . . . . .	22
3.	Familias de cóputas . . . . .	23
3.1.	Cóputas elípticas . . . . .	23
3.2.	Cóputas arquimedianas . . . . .	26
3.3.	Cóputa bivariada de Marshall Olkin . . . . .	29
3.4.	Conclusiones . . . . .	31
4.	Medidas de dependencia . . . . .	31
4.1.	Correlación lineal . . . . .	32
4.2.	Correlación de rangos . . . . .	33
4.3.	Resumen del capítulo . . . . .	35
5.	Funciones y coeficientes de dependencia . . . . .	35
5.1.	Coficientes de dependencia de cola . . . . .	35
5.2.	Resultados generales . . . . .	41
5.3.	Funciones y coeficientes de dependencia generales . . . . .	43
5.4.	Resumen del capítulo . . . . .	45
6.	Estimación y ajuste de cóputas y funciones de dependencia extrema . . . . .	45
6.1.	Cóputa empírica . . . . .	46
6.2.	Estimación de los parámetros de la cóputa por el método de Máxima verosimilitud . . . . .	46
6.3.	Bondad de ajuste: Crámer von Mises (CvM) . . . . .	47
6.4.	Estimación no paramétrica de la función de dependencia extrema . . . . .	48
6.5.	Resumen del capítulo . . . . .	49

7.	Simulación de cópulas . . . . .	50
7.1.	Cópula Gaussiana y t-Cópula . . . . .	50
7.2.	Cópulas arquimedianas . . . . .	51
7.3.	Cópula de Marshall-Olkin . . . . .	51
7.4.	Cópulas en R . . . . .	52
7.5.	Resumen del capítulo . . . . .	53
8.	Fundamentos del Análisis de series de tiempo financieras . . . . .	53
8.1.	Series de tiempo financieras . . . . .	53
8.2.	Hechos estilizados en datos financieros . . . . .	62
8.3.	Resumen del capítulo . . . . .	63
9.	Cálculo del VaR utilizando funciones de dependencia extrema y cópulas . . . . .	63
9.1.	Introducción . . . . .	65
9.2.	Valor en riesgo (VaR) . . . . .	67
9.3.	Cálculo del VaR de un portafolio . . . . .	67
9.4.	Descripción de las series . . . . .	68
9.5.	Ajuste del modelo ARIMA-GARCH . . . . .	70
9.6.	Ajuste de la distribución Pareto Generalizada y cópula . . . . .	80
9.7.	Simulaciones finales . . . . .	86
9.8.	Cálculo del VaR . . . . .	87
9.9.	Resumen del capítulo . . . . .	88
10.	Conclusiones generales . . . . .	88
11.	Apéndice . . . . .	91

# Agradecimientos

*Quiero agradecer a Dios por la grandiosa oportunidad que me ha dado al formar parte de este instante en la historia de la Tierra.*

*Mención especial merece mi familia, particularmente: Mi madre por ser mi compañera incondicional, mi apoyo, mi mano derecha, mi todo. Mi padre por estar ahí siempre que lo he necesitado y por su paciencia. Y mi hermano, por protegerme y cuidarme, por ser un ejemplo para mí. A los tres, gracias. Nada de lo que he hecho hubiera sido posible sin ustedes.*

*A mi asesor, el Dr. Yuri Salazar, por su tiempo, paciencia, ayuda, entrega, y por no permitirme conformarme.*

*A mis sinodales: Dra. Ana Meda, Dr. Adán Díaz, M. en C. Laura Liljehult y M en C. Daniel Cervantes por sus útiles y oportunos comentarios a esta tesis, así como por su tiempo y su paciencia.*

*A mi Universidad Nacional Autónoma de México, por darme las herramientas para forjar mi futuro así como por regalarme algunos de los mejores momentos de mi vida.*

## 1. Introducción

En probabilidad y estadística siempre ha sido importante conocer la relación que guardan las funciones de distribución multivariadas y sus correspondientes funciones de distribución univariadas. Diversos autores trabajaron en el tema, pero fue Abe Sklar quien en 1959 formalizó el concepto de cópula, y además enunció el teorema que hoy lleva su apellido y que nos demuestra que si dos (o más) variables aleatorias cumplen con los requisitos expuestos en el hoy conocido como Teorema de Sklar, existe una única cópula que junto con las funciones de distribución marginales, logran reconstruir la función de distribución conjunta.

El estudio de las cópulas es una alternativa a las medidas de dependencia ya conocidas, por ejemplo la correlación lineal (Pearson) que, como su nombre lo indica, se limita a relaciones lineales. La correlación nos da una manera de cuantificar la dependencia lineal entre dos variables, aunque un valor de cero podría representar independencia solo si las variables tienen una estructura de dependencia elíptica. Los coeficientes de correlación de rangos miden otras formas de dependencia (no sólo lineal), sin embargo, son una medida cuantitativa y no proveen mayor información sobre la relación entre las variables.

Por lo anterior, una manera de conocer más información acerca de la estructura de dependencia es utilizar cópulas, ya que éstas cuentan con la ventaja de que no se encuentran limitadas a relaciones lineales. Una característica muy importante de las cópulas es que capturan la dependencia entre  $n$  variables independientemente de cuáles sean las funciones de distribución marginales. Más aún, las cópulas nos dan un enfoque para la construcción de modelos multivariados. La cópula nos permite combinar nuestros modelos marginales más complejos con una gran variedad de modelos de dependencia.

Además, con ellas seremos capaces de medir la dependencia para distribuciones con colas pesadas. Lo anterior es muy importante en la medición de riesgos pues los casos en las colas de las distribuciones suelen estar asociados a pérdidas muy grandes (o ganancias)).

Como veremos más adelante, el conocer la función de distribución multivariada nos permite reconstruir a las marginales, pero el hecho de conocer únicamente a las marginales no nos permite conocer la distribución conjunta multivariada. Las cópulas son las funciones que nos hacen falta para que, junto con las funciones de distribución univariadas, logremos reconstruir las funciones de distribución multivariadas.

La estructura de esta tesis es la siguiente: En el capítulo 1, comenzaremos con algunos resultados preliminares que nos servirán como repaso y ayudarán a tener una mejor comprensión de los temas siguientes. Continuaremos con una definición de cópulas en términos de funciones de distribución y enunciaremos el Teorema de Sklar. Asimismo, revisaremos las cotas que tienen todas las cópulas y los ejemplos más representativos que existen de cópulas.

En el segundo capítulo, revisaremos las familias de cópulas elípticas, arquimedianas y la cópula de Marshall-Olkin. En el tercer capítulo analizaremos dos medidas de dependencia comúnmente utilizados: la correlación lineal y la correlación de rangos. Al finalizar el capítulo, revisaremos los coeficientes de dependencia de cola, y para su definición, utilizaremos la definición de la cópula.

En el cuarto capítulo revisaremos las funciones y los coeficientes de dependencia, así como algunas de sus características. Abordaremos el concepto de cópula de supervivencia y simetría, finalizando con la generalización de los coeficientes de dependencia extrema con la ayuda de las llamadas, funciones de dependencia extrema.

En el quinto capítulo revisaremos la definición de cópula empírica, indispensable para la estimación de parámetros por máxima verosimilitud y las pruebas de bondad de ajuste, mismas que se revisarán también en este capítulo. Se mostrará una propuesta para la función de dependencia extrema. En el sexto capítulo, presentaremos cómo se realiza la simulación de cópulas y al final del mismo, revisaremos como se implementan estas simulaciones en el software estadístico R, para algunas cópulas.

En el séptimo capítulo se realizará un breve recuento de los fundamentos para el análisis de series de tiempo así como de los hechos estilizados presentes en las series financieras. Finalmente, el último capítulo aborda una situación práctica en la que se utilizarán las cópulas para modelar la dependencia entre las series de residuales de dos acciones que conforman un portafolio teórico y del cual evaluaremos su rendimiento futuro a corto plazo. Para validar y calibrar el modelo, realizaremos *backtesting* para el primer mes del 2016 y compararemos el comportamiento real del portafolio contra el simulado.

## 2. Preliminares

Al trabajar con cópulas se debe estar familiarizado con las operaciones de probabilidad, transformaciones de cuantiles y con las propiedades de la función inversa generalizada. A continuación, mencionaremos de manera breve algunos resultados que serán utilizados en el desarrollo de esta tesis.

Considere la siguiente notación:

- $\mathbb{R}$  denota la recta real  $(-\infty, \infty)$
- $\bar{\mathbb{R}}$  denota la recta real extendida  $[-\infty, \infty]$
- $\bar{\mathbb{R}}^2$  es el plano real extendido  $\bar{\mathbb{R}} \times \bar{\mathbb{R}}$



- Un rectángulo en  $\bar{\mathbb{R}}^2$  es el producto cartesiano  $B$  de dos intervalos cerrados  $B = [x_1, x_2] \times [y_1, y_2]$ . Los vértices del rectángulo son los puntos:  $(x_1, y_1)$ ,  $(x_1, y_2)$ ,  $(x_2, y_1)$  y  $(x_2, y_2)$
- El cuadrado unitario  $I^2$  es el producto  $I \times I$ , donde  $I = [0, 1]$ .

Sea  $T$  una función creciente, es decir, una función que satisface que  $y > x \Rightarrow T(y) \geq T(x)$  con desigualdad estricta en el lado izquierdo para alguna pareja  $y > x$ . Así, una función creciente puede tener secciones planas, si queremos evitar esto, estipularemos entonces que  $T$  sea estrictamente creciente, esto es  $y > x \Rightarrow T(y) > T(x)$ .

**Lema 1.** 1. Si  $X$  es una variable aleatoria y  $T$  es una función creciente, entonces

$$\{X \leq x\} \subset \{T(X) \leq T(x)\}$$

y

$$\mathbb{P}(T(X) \leq T(x)) = P(X \leq x) + P(T(X) = T(x), X > x)$$

2. Si  $F$  es la función de distribución de  $X$ , entonces:

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

**Definición 1.** Sea  $F$  una función de distribución. Luego, la cuasi-inversa de  $F$  (o la inversa generalizada de  $F$ ) es una función  $F^{\leftarrow}(t)$  con dominio  $I=[0,1]$  tal que:

1. Si  $t$  está en el  $\text{Ran}F$ , entonces  $F^{\leftarrow}(t)$  es cualquier número  $x$  en  $\bar{\mathbb{R}} = [-\infty, \infty]$  tal que  $F(x) = t$ , es decir, para toda  $t$  en  $\text{Ran}F$ ,  $F(F^{\leftarrow}(t)) = t$ ;
2. Si  $t$  no está en el  $\text{Ran}F$ , entonces:

$$F^{\leftarrow}(t) = \inf \{x \mid F(x) \geq t\} = \sup \{x \mid F(x) \leq t\}$$

Si  $F$  es estrictamente creciente, entonces tiene una única cuasi-inversa, que es la inversa ordinaria ( $F^{-1}$ ).

**Proposición 1.** (Propiedades de la inversa generalizada) Sea  $T$  una función creciente i.e. una función que satisface  $y > x \Rightarrow T(y) \geq T(x)$  con la desigualdad estricta en el lado derecho para alguna pareja  $y > x$ . Se cumple lo siguiente:

1.  $T^{\leftarrow}$  es una función creciente y continua por la izquierda
2.  $T$  es continua  $\Leftrightarrow T^{\leftarrow}$  es estrictamente creciente.
3.  $T$  es estrictamente creciente  $\Leftrightarrow T^{\leftarrow}$  es continua.

Para las siguientes propiedades considere además que  $T^{\leftarrow} < \infty$ :

4. Si  $T$  es una función continua por la derecha,  $T(x) > y \Leftrightarrow T^{\leftarrow}(y) \leq x$
5.  $T^{\leftarrow} \circ T(x) \leq x$
6.  $T \circ T^{\leftarrow}(y) \geq y$
7.  $T$  es estrictamente creciente  $\Rightarrow T^{\leftarrow} \circ T(x) = x$
8.  $T$  es continua  $\Rightarrow T \circ T^{\leftarrow}(y) = y$ .

Aplicaremos la idea de la inversa generalizada a las funciones de distribución. Si  $F$  es una función de distribución, luego la inversa generalizada  $F^{\leftarrow}$  se conoce como la *función cuantil* de  $F$ . En este caso, para  $\alpha \in (0, 1)$ , también usaremos la notación  $q_\alpha(F) = F^{\leftarrow}(\alpha)$  para el  $\alpha$ -cuantil de  $F$ .

En general, debido a que las funciones de distribución no necesariamente son estrictamente crecientes, tenemos que  $F^{\leftarrow} \circ F(x) \leq x$ , por Proposición 1 (5). Pero los valores de  $x$ , donde  $F^{\leftarrow} \circ F(x) \neq x$ , corresponde a piezas planas y tienen masa de probabilidad igual a cero. Por lo tanto, tenemos el siguiente hecho.

**Proposición 2.** Si  $X$  es una variable aleatoria con función de distribución  $F$ , entonces  $\mathbb{P}(F^{\leftarrow} \circ F(X) = X) = 1$

**Proposición 3.** Sea  $F$  una función de distribución y sea  $F^{\leftarrow}$  su inversa generalizada, es decir, la función  $F^{\leftarrow}(y) := \inf\{x : F(x) \geq y\}$ :

1. *Transformación cuantil*  
Si  $U$  tiene distribución uniforme estándar, entonces  $\mathbb{P}(F^{\leftarrow}(U) \leq x) = F(x)$ .
2. *Transformación de probabilidad*  
Si  $X_2$  tiene función de distribución  $F$ , donde  $F$  es una función de distribución univariada continua, entonces  $F(X_2) \sim U(0, 1)$

*Demostración.* Sea  $x_2 \in \mathbb{R}$  y  $u \in (0, 1)$ . Para demostrar el primer punto, recuerde que:

$$F(x_2) \geq u \Leftrightarrow F^{\leftarrow}(u) \leq x_2$$

(véase Proposición 1(4)), por lo que se sigue que:

$$\mathbb{P}(F^{\leftarrow}(U) \leq x_2) = \mathbb{P}(U \leq F(x_2)) = F(x_2)$$

Para la segunda parte de la demostración:

$$\begin{aligned} \mathbb{P}(F(X_2) \leq u) &= \mathbb{P}(F^{\leftarrow} \circ F(X_2) \leq F^{\leftarrow}(u)) \\ &= \mathbb{P}(X_2 \leq F^{\leftarrow}(u)) \\ &= F \circ F^{\leftarrow}(u) \\ &= u \end{aligned}$$

Donde la primera igualdad se sigue del hecho de que  $F^{\leftarrow}$  es estrictamente creciente (Proposición 1 (2)), la segunda se sigue de Proposición 2, y la última de Proposición 1(8).  $\square$

La primera parte de la proposición anterior es clave en la simulación estocástica. Si podemos generar una distribución uniforme y calcular la inversa de la distribución  $F$ , entonces podemos generar una muestra de esa distribución. Ambas partes de la proposición implican que podemos transformar variables con una función de distribución particular continua a otra función de distribución continua. Una de las principales definiciones que se estudia cuando se inicia el estudio de vectores aleatorios es la siguiente:

**Definición 2.** (*Función de distribución marginal*)

Sea  $(X_1, X_2)$  un vector aleatorio con función de distribución  $F(x_1, x_2)$ . A la función

$$F_1(x_1) = \lim_{x_2 \rightarrow \infty} F(x_1, x_2)$$

se le conoce como la función de distribución marginal de  $X_1$ . Análogamente se define la función de distribución marginal de  $X_2$  como

$$F_2(x_2) = \lim_{x_1 \rightarrow \infty} F(x_1, x_2)$$

Ilustremos lo anterior con un ejemplo. Sea

$$F(x_1, x_2) = \begin{cases} 0 & , \text{ si } x_1 < 0 \text{ o } x_2 < 0, \\ x_1 x_2 / 4 & , \text{ si } 0 \leq x_1, x_2 < 2, \\ x_1 / 2 & , \text{ si } 0 \leq x_1 < 2, x_2 \geq 2, \\ x_2 / 2 & , \text{ si } 0 \leq x_2 < 2, x_1 \geq 2, \\ 1 & , \text{ si } x_1 \geq 2 \text{ y } x_2 \geq 2 \end{cases}$$

Para encontrar, por ejemplo, la función de distribución marginal de  $X_1$  lo único que debemos hacer es calcular el límite de  $F(x_1, x_2)$  cuando  $x_2$  tiende a  $\infty$ , lo cual da como resultado:

$$F_1(x_1) = \begin{cases} 0 & , \text{ si } x_1 < 0, \\ x_1 / 2 & , \text{ si } 0 \leq x_1 < 2, \\ 1 & , \text{ si } x_1 \geq 2, \end{cases}$$

Y de manera análoga, haciendo tender  $x_1$  a  $\infty$  podemos obtener la función de distribución marginal de  $X_2$ :

$$F_2(x_2) = \begin{cases} 0 & , \text{ si } x_2 < 0, \\ x_2 / 2 & , \text{ si } 0 \leq x_2 < 2, \\ 1 & , \text{ si } x_2 \geq 2, \end{cases}$$

Con lo anterior, nos percatamos de que si conocemos la función de distribución conjunta  $F(x_1, x_2)$  podemos conocer las funciones de distribución marginales de  $X_1$  y  $X_2$ ,  $F_1(x_1)$  y  $F_2(x_2)$ , respectivamente. Dicho esto, es natural preguntarse si el recíproco de la afirmación

anterior también es cierto, es decir, dado que conocemos las funciones de distribución marginales  $F_1(x_1)$  y  $F_2(x_2)$ , ¿Podemos conocer la función de distribución conjunta  $F(x_1, x_2)$ ?

Para responder a esta interrogante, veamos el siguiente ejemplo:

Sean  $f_1$  y  $f_2$  funciones de densidad de probabilidades y  $F_1(x_1)$ ,  $F_2(x_2)$  sus respectivas funciones de distribución marginales. Definimos a la familia paramétrica de funciones bivariadas  $f_\lambda : \mathbb{R}^2 \rightarrow \mathbb{R}$  como:

$$f_\lambda(x_1, x_2) = f_1(x_1)f_2(x_2)[1 + \lambda(1 - 2F_1(x_1))(1 - 2F_2(x_2))]$$

Es fácil verificar que cuando  $-1 \leq \lambda \leq 1$ ,  $f_\lambda$  es una familia de funciones de densidad conjunta con marginales  $f_1$  y  $f_2$ . Demostraremos que es función de densidad conjunta:

$$-1 \leq 1 - 2F_1(x_1) \leq 1 \text{ y } -1 \leq 1 - 2F_2(x_2) \leq 1$$

y como  $-1 \leq \lambda \leq 1$

$$-1 \leq \lambda(1 - 2F_1(x_1))(1 - 2F_2(x_2)) \leq 1$$

con lo que concluimos que  $f_\lambda(x_1, x_2) \geq 0$  para todo punto  $(x_1, x_2) \in \mathbb{R}^2$  y también,

$$\begin{aligned} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_\lambda dx_1 dx_2 &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_1(x_1)f_2(x_2) dx_1 dx_2 + \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \lambda f_1(x_1)f_2(x_2) dx_1 dx_2 \\ &\quad - 2\lambda \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} F_2(x_2)f_1(x_1)f_2(x_2) dx_1 dx_2 \\ &\quad - 2\lambda \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} F_1(x_1)f_1(x_1)f_2(x_2) dx_1 dx_2 \\ &\quad + 4\lambda \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} F_1(x_1)F_2(x_2)f_1(x_1)f_2(x_2) dx_1 dx_2 \\ &= 1 + \lambda - 2\lambda \frac{F_1^2(x_1)}{2} \Big|_{-\infty}^{\infty} - 2\lambda \frac{F_2^2(x_2)}{2} \Big|_{-\infty}^{\infty} + 4\frac{\lambda}{4} \\ &= 1 + \lambda - \lambda - \lambda + \lambda \\ &= 1 \end{aligned}$$

En efecto,  $f_\lambda$  es una función de densidad conjunta. Veamos ahora quiénes son las funciones de distribución marginales:

$$\begin{aligned}
f_1(x_1) &= \int_{-\infty}^{\infty} f_{\lambda} dx_2 = \int_{-\infty}^{\infty} f_1(x_1)f_2(x_2)dx_2 + \int_{-\infty}^{\infty} \lambda f_1(x_1)f_2(x_2)dx_2 \\
&\quad - 2\lambda \int_{-\infty}^{\infty} F_2(x_2)f_1(x_1)f_2(x_2)dx_2 \\
&\quad - 2\lambda \int_{-\infty}^{\infty} F_1(x_2)f_1(x_1)f_2(x_2)dx_2 \\
&\quad + 4\lambda \int_{-\infty}^{\infty} F_1(x_1)F_2(x_2)f_1(x_1)f_2(x_2)dx_2 \\
&= f_1(x_1) + \lambda f_1(x_1) - 2\lambda F_1(x_1)f_1(x_1) - \lambda f_1(x_1) + 2\lambda F_1 f_1 = f_1(x_1)
\end{aligned}$$

De manera análoga, se obtiene que la otra marginal es  $f_2(x_2)$ . Este es un claro ejemplo de que, en general, cuando únicamente se conocen las funciones de distribución marginales ésta información no es suficiente para determinar la función de distribución conjunta.

Para definir qué es una cópula, revisaremos antes las definiciones de funciones ancladas y 2-crecientes:

**Definición 3.** Sean  $S_1, S_2$  dos conjuntos no vacíos de  $\bar{\mathbb{R}}$ , y sea  $H$  una función bidimensional tal que el  $\text{Dom}H = S_1 \times S_2$ . Sea  $B = [x_1, x_2] \times [y_1, y_2]$  un rectángulo cuyos vértices están en  $\text{Dom}H$ , entonces el  $H$ -volumen de  $B$  está dado por:

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

**Definición 4.** Una función real es 2-creciente si  $V_H(B) \geq 0$  para todos los rectángulos  $B$  cuyos vértices estén en  $\text{Dom}H$ .

**Definición 5.** Sea  $F$  una función tal que  $F : I^2 \rightarrow I$ . Decimos que  $F$  es una función anclada si cumple que para cada  $u, v$  en  $I$ ,  $C(u, 0) = 0 = C(0, v)$ .

## 2.1. Cópulas

Entonces, ¿Qué nos hace falta para poder determinar la función de distribución conjunta de  $X_1$  y  $X_2$ , dado que conocemos las marginales? Si regresamos al ejemplo anterior, nos percatamos de que aún cuando conocíamos las marginales no conocíamos la *manera* en la que éstas se asociaban. Es decir, para poder determinar a la función de distribución conjunta necesitamos no solamente de las marginales sino también de una función que describa la manera en la que éstas se interrelacionan. A ésta función se le conoce como *cópula* y a continuación enunciaremos su definición formal:

**Definición 6.** Una cópula bidimensional es una función  $C : I \times I \mapsto I$  con las siguientes propiedades:

1. Para cada  $u, v$  en  $I$ :  $C(u, v)$  es creciente en cada componente
2. Para cada  $u, v$  en  $I$ :  $C(u, 0) = 0 = C(0, v)$   $C(u, 1) = u$  y  $C(1, v) = v$
3. Para cada  $u_1, u_2, v_1, v_2$  en  $I$  tal que  $u_1 \leq u_2$  y  $v_1 \leq v_2$ ,

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

La primera y la tercera propiedad son un requisito para cualquier función de distribución bivariada, mientras que la segunda es un requisito de marginales uniformes. A la tercera se le conoce como desigualdad del rectángulo y establece que si el vector aleatorio  $(U_1, U_2)$  tiene función de distribución  $C$  entonces,  $\mathbb{P}(a_1 \leq U_1 \leq b_1, a_2 \leq U_2 \leq b_2)$  es no negativa.

De manera alternativa, podemos definir a la cópula como sigue:

**Definición 7.** Una función  $C : I^2 \mapsto I$  es una cópula si es una función de distribución correspondiente a un vector aleatorio  $(U, V)$ , donde  $U \sim (0, 1)$  y  $V \sim (0, 1)$ .

## 2.2. Teorema de Sklar

Es momento de presentar el teorema central de la teoría de cópulas, el llamado *Teorema de Sklar*. En 1959, Abe Sklar explicó el papel que juegan las cópulas en la relación entre las funciones de distribución conjunta multivariadas y sus marginales unidimensionales en el teorema que hoy lleva su apellido. En él, expresa a la función de distribución conjunta en función de la cópulas y las funciones de distribución marginales. El teorema puede ser invertido, es decir, se puede expresar a la cópula en términos de la función de distribución conjunta y sus marginales, lo que nos permite utilizar a las cópulas de manera conjunta con las marginales para construir nuevas funciones de distribución multivariadas.

**Teorema 1.** (*Teorema de Sklar*) Sea  $F$  una función de distribución conjunta con marginales  $F_1$  y  $F_2$ . Entonces, existe una cópula  $C$  tal que para toda pareja  $(x_1, x_2)$  en  $\bar{\mathbb{R}}^2$  con  $\bar{\mathbb{R}} = [-\infty, \infty]$ ,

$$F(x_1, x_2) = C(F_1(x_1), F_2(x_2)) \tag{1}$$

Si  $F_1$  y  $F_2$  son continuas, entonces  $C$  es única; de otra manera,  $C$  sólo se determina de manera única sobre  $\text{Ran}F_1 \times \text{Ran}F_2$ . Por otro lado, si  $C$  es una cópula y  $F_1$  y  $F_2$  son funciones de distribución marginales, entonces la función  $F$  definida arriba es una función de distribución conjunta con marginales  $F_1$  y  $F_2$ .

*Demostración.* Probaremos la existencia y unicidad de la cópula en el caso cuando  $F_1, F_2$  son continuas y la segunda parte del teorema en su forma general. Para una prueba completa vea **Nelsen (1999)**.

Para cualquier  $x_1, x_2$  en  $\bar{\mathbb{R}}$  inferiremos que si  $(X_1, X_2)$  tiene función de distribución  $F$  entonces:

$$F(X_1, X_2) = \mathbb{P}(F_1(X_1) \leq F_1(x_1), F_2(X_2) \leq F_2(x_2))$$

Debido a que  $F_1, F_2$  son continuas, la Proposición 1(2) y la Definición 6 implican que la función de distribución de  $(F_1(X_1), F_2(X_2))$  es una cópula, que denotaremos por  $C$ , y entonces tenemos la primera parte del teorema.

Si evaluamos la igualdad de la primera parte del teorema en los argumentos  $x_i = F_i^{\leftarrow}(u_i), 0 \leq u_i \leq 1, i = 1, 2$  y usamos la Proposición 1(3), obtenemos:

$$C(u_1, u_2) = F(F_1^{\leftarrow}(u_1), F_2^{\leftarrow}(u_2))$$

lo cual nos proporciona una representación explícita de  $C$  en términos de  $F$  y de sus marginales, y con ello demostramos unicidad.

Para la última parte del teorema, suponga que  $C$  es una cópula y que  $F_1, F_2$  son funciones de distribución univariadas. Construimos un vector aleatorio con la distribución dada en la ecuación del teorema de Sklar, tomando  $U = (U_1, U_2)$  como el vector aleatorio con distribución  $C$  y definiendo:  $X := (F_1^{\leftarrow}(U_1), F_2^{\leftarrow}(U_2))$ .

Luego verificamos, usando la Proposición 1(8), que:

$$\begin{aligned} \mathbb{P}(X_1 \leq x_1, X_2 \leq x_2) &= \mathbb{P}(F_1^{\leftarrow}(U_1) \leq x_1, F_2^{\leftarrow}(U_2) \leq x_2) \\ &= \mathbb{P}(U_1 \leq F_1(x_1), U_2 \leq F_2(x_2)) \\ &= C(F_1(x_1), F_2(x_2)) \end{aligned}$$

□

La ecuación 1 nos da una expresión para las funciones de distribución conjuntas en términos de una cópula y dos funciones de distribución univariadas. Sin embargo, esta ecuación puede ser invertida para expresar a las cópulas en términos de la función de distribución conjunta y de las *inversas* de las dos marginales.

Debemos notar que si una marginal no es estrictamente creciente, entonces no posee inversa en el sentido usual, es por lo anterior que utilizaremos la definición de cuasi-inversas. Utilizando la Definición 1, tenemos el siguiente corolario:

**Corolario 1.** Sean  $X_1, X_2$  variables aleatorias con  $F_1, F_2$  sus respectivas funciones de distribución continuas,  $F$  la función de distribución conjunta,  $C$  la cópula de  $X, Y$  y sean  $F_1^{(-1)}$  y  $F_2^{(-1)}$  cuasi-inversas de  $F_1$  y  $F_2$  respectivamente. Entonces para cualquier  $(u, v)$  en  $DomC$ :

$$C(u, v) = F(F_1^{(-1)}(u), F_2^{(-1)}(v)) \quad (2)$$

Las expresiones 1 y 2 son fundamentales cuando se trabaja con cópulas. La primera demuestra como las funciones de distribución conjuntas  $F$  son formadas al ser acopladas junto a sus distribuciones marginales con cópula  $C$ , la segunda muestra como las cópulas son extraídas de las funciones de distribución multivariadas con marginales continuas.

**Definición 8.** Si el vector aleatorio  $X$  tiene función de distribución  $F$  con funciones de distribución marginales  $F_1$  y  $F_2$ , entonces la cópula de  $X$  es la función de distribución  $C$  de  $(F_1(X_1), F_2(X_2))$ .

Veamos un ejemplo:

(Distribución Gumbel exponencial bivariada.) Sea  $F_\lambda$  una función de distribución conjunta dada por:

$$F_\lambda(x_1, x_2) = \begin{cases} 1 - e^{-x_1} - e^{-x_2} + e^{-(x_1+x_2+\lambda x_1 x_2)}, & \text{si } x_1 \leq 0, x_2 \leq 0, \\ 0, & \text{si c.o.c} \end{cases}$$

Donde  $\lambda$  es un parámetro en  $I$  y las funciones de distribución marginales son exponenciales con cuasi-inversas  $F_1^{(-1)}(u) = -\ln(1-u)$  y  $F_2^{(-1)}(v) = -\ln(1-v)$  para  $u, v$  en  $I$ . Así, la cópula correspondiente es:

$$C_\lambda(u, v) = F(F_1^{(-1)}(u), F_2^{(-1)}(v)) = u + v - 1 + (1-u)(1-v)e^{-\lambda \ln(1-u)\ln(1-v)}$$

Como podemos apreciar en este ejemplo, si conocemos la función de distribución conjunta  $F$ , podemos obtener las marginales  $F_1$  y  $F_2$  y con esto, conocer la cópula  $C$  con la ayuda del Teorema 1 y el Corolario 1.

Pensemos ahora en que únicamente conocemos  $C$ , ¿Podemos con esta información reconstruir a  $F$ ? La respuesta también está dada por el teorema de Sklar ya que en él se demuestra que existe una función  $C$  que establece la relación funcional entre la distribución conjunta y sus marginales, esto es, aunque con  $C$  conocemos la relación de dependencia entre la función de distribución y sus marginales, de ella no podemos obtener información alguna de las marginales por lo que no tenemos la información necesaria para construir a  $F$ .

Sea  $F(x_1, x_2)$  una función de distribución y supongamos que  $F(x_1, x_2)$  es normal bivariada. Analicemos con cuidado las implicaciones que existen detrás de hacer esta suposición, al ser  $F$  normal bivariada suponemos que sus marginales  $F_1$  y  $F_2$  son también normales y además suponemos que la relación entre ellas es también normal. En otras palabras, al suponer una distribución multivariada predisponemos a las marginales y a la relación entre ellas a seguir una distribución determinada.

El utilizar cópulas nos permite resolver este problema ya que podemos modelar a las marginales de manera separada a la estructura de dependencia. Lo anterior nos brinda mayor flexibilidad para modelar, ya que podemos modelar a cada marginal por separado y después



introducir la estructura de dependencia con una cópula apropiada.

Notemos que si tenemos una función de distribución conjunta bivariada  $F$ , con funciones de distribución marginales continuas  $F_1, F_2$  podemos extraer la cópula subyacente  $C$  mediante el Corolario 1 y después construir una nueva función de distribución conjunta de probabilidades bivariada  $F^*$  con la misma cópula pero con distintas funciones de distribución marginales  $F_1^*$  y  $F_2^*$ , esto es:

$$F^*(x_1, x_2) = C(F_1^*, F_2^*)$$

En varios textos de estadística se pide encontrar un ejemplo de una distribución bivariada con marginales normal estándar que no sea normal con parámetros  $\mu_x = \mu_y = 0$  y  $\sigma_x^2 = \sigma_y^2 = 1$  y el coeficiente de Pearson  $\rho$ . Responder al ejercicio se vuelve fácil, gracias al teorema de Sklar. Sea  $C$  la cópula tal que:

$$C_\theta(u, v) = u + v - 1 + (1 - u)(1 - v)e^{-\theta \ln(1-u)\ln(1-v)}$$

y combinándola con las marginales normal estándar en la fórmula  $H(x, y) = C(F(x), G(y))$ , tenemos una respuesta al problema planteado. De hecho, si  $\Phi$  denota la función de distribución normal estándar univariada y  $N_\rho$  denota la función de distribución normal estándar bivariada con  $\rho$ , entonces cualquier cópula excepto una de la forma:

$$C(u, v) = N_\rho(\Phi^{-1}(u), \Phi^{-1}(v)) = \frac{1}{2\pi\sqrt{1-\rho^2}} \int_{-\infty}^{\Phi^{-1}(u)} \int_{-\infty}^{\Phi^{-1}(v)} e^{-\frac{(s^2-2\rho st+t^2)}{2(1-\rho^2)}} ds dt$$

con  $\rho \neq 1, 0, -1$  cumple con lo solicitado.

Como ya veíamos en el párrafo pasado, el teorema de Sklar es muy importante ya que provee un método para la construcción de funciones de distribución conjunta. Dicha construcción consiste en lo siguiente: dadas una función de distribución continua y sus marginales, podemos calcular la cópula mediante el Teorema 1. Con ello, lo que hicimos fue extraer la estructura de dependencia entre dos variables y si *alimentamos* a  $C$  con marginales distintas obtenemos una función de distribución distinta con la misma estructura de dependencia.

Es importante no perder de vista que el concepto de cópula es menos natural en las distribuciones discretas bivariadas. Esto se debe a que más de una cópula puede ser usada para unir las marginales y formar la función de distribución conjunta, como veremos en el siguiente ejemplo:

**Ejemplo: Cópula de una Bernoulli bivariada**

Sea  $(X_1, X_2)$  un vector aleatorio con función de distribución Bernoulli que satisfice:

$$\begin{aligned} \mathbb{P}(X_1 = 0, X_2 = 0) &= \frac{1}{8}, \\ \mathbb{P}(X_1 = 0, X_2 = 1) &= \frac{7}{8}, \end{aligned}$$

$$\begin{aligned}\mathbb{P}(X_1 = 1, X_2 = 0) &= \frac{2}{8}, \\ \mathbb{P}(X_1 = 1, X_2 = 1) &= \frac{1}{8}\end{aligned}$$

Claramente,  $\mathbb{P}(X_1 = 0) = \mathbb{P}(X_2 = 0) = \frac{3}{8}$  y las distribuciones marginales  $F_1$  y  $F_2$  de  $X_1$  y  $X_2$  son las mismas. Por el Teorema de Sklar, sabemos que:

$$\mathbb{P}(X_1 \leq x_1, X_2 \leq x_2) = C(\mathbb{P}(X_1 \leq x_1), \mathbb{P}(X_2 \leq x_2))$$

para toda  $x_1, x_2$  y alguna cópula  $C$ . Como el  $\text{Ran } F_1 = \text{Ran } F_2 = \{0, \frac{3}{8}, 1\}$ , tenemos una restricción sobre  $C$ :  $C(\frac{3}{8}, \frac{3}{8}) = \frac{1}{8}$ . Cualquier cópula que satisfaga esta condición es una cópula de  $(X_1, X_2)$ , y hay una infinidad de posibilidades.

*Invarianza.* Una propiedad poderosa de la cópula de una distribución es su invarianza bajo transformaciones estrictamente crecientes de las marginales. Bajo el punto de vista del teorema de Sklar y esta propiedad de invarianza, interpretamos a la cópula de una distribución como una manera muy natural de representar la estructura de dependencia de esa distribución, aunque sólo en el caso de marginales continuas.

**Teorema 2.** Sean  $X_1$  y  $X_2$  variables aleatorias continuas con cópula  $C_{XY}$ . Si  $\alpha$  y  $\beta$  son estrictamente crecientes sobre el  $\text{Ran } X$  y  $\text{Ran } Y$  respectivamente, entonces la cópula  $C_{\alpha(X), \beta(Y)} = C_{XY}$ . Así,  $C_{XY}$  es invariante bajo transformaciones estrictamente crecientes de  $X$  y  $Y$ .

*Demostración.* Sean  $F_1, G_1, F_2$  y  $G_2$  funciones de distribución de  $X_1, X_2, \alpha(X_1), \beta(X_2)$  respectivamente. Como  $\alpha$  y  $\beta$  son estrictamente crecientes tenemos que:

$$F_2(x_1) = \mathbb{P}(\alpha(X_1) \leq x_1) = \mathbb{P}(X_1 \leq \alpha^{-1}(x_1)) = F_1(\alpha^{-1}(x_1))$$

Y también:

$$G_2(x_2) = \mathbb{P}(\beta(X_2) \leq x_2) = \mathbb{P}(X_2 \leq \beta^{-1}(x_2)) = G_1(\beta^{-1}(x_2))$$

Así, para cualquier  $x_1, x_2$ , tenemos que:

$$\begin{aligned}C_{\alpha(X_1), \beta(X_2)}(F_2(x_1), G_2(x_2)) &= \mathbb{P}(\alpha(X_1) \leq x_1, \beta(X_2) \leq x_2) \\ &= \mathbb{P}(X_1 \leq \alpha^{-1}(x_1), X_2 \leq \beta^{-1}(x_2)) \\ &= C_{X_1 X_2}(F_1(\alpha^{-1}(x_1)), G_1(\beta^{-1}(x_2))) \\ &= C_{X_1 X_2}(F_2(x_1), G_2(x_2))\end{aligned}$$

Como  $X_1$  y  $X_2$  son continuas,  $\text{Ran } F_2 = \text{Ran } G_2 = I$  por lo que se sigue que  $C_{\alpha(X_1), \beta(X_2)} = C_{X_1 X_2}$  en  $[0, 1] \times [0, 1]$ .  $\square$

Cuando al menos  $\alpha$  o  $\beta$  son estrictamente crecientes, obtenemos resultados en los cuales la cópula de las variables aleatorias  $\alpha(X_1)$  y  $\beta(X_2)$  es una simple transformación de  $C_{X_1 X_2}$ . De manera específica, tenemos que:

**Teorema 3.** Sean  $X_1, X_2$  variables aleatorias continuas con cópula  $C_{X_1 X_2}$ . Sea  $\alpha, \beta$  estrictamente monótonas sobre  $\text{Ran } X_1, \text{Ran } X_2$  respectivamente, tenemos que:

- Si  $\alpha$  es estrictamente creciente y  $\beta$  es estrictamente decreciente, entonces:

$$C_{\alpha(X_1),\beta(X_2)}(u, v) = u - C_{X_1X_2}(u, 1 - v)$$

- Si  $\alpha$  es estrictamente decreciente y  $\beta$  es estrictamente creciente, entonces:

$$C_{\alpha(X_1),\beta(X_2)}(u, v) = v - C_{X_1X_2}(1 - u, v)$$

- Si  $\alpha$  y  $\beta$  son ambas estrictamente decrecientes, entonces:

$$C_{\alpha(X_1),\beta(X_1)}(u, v) = u + v - 1 + C_{X_1X_2}(1 - u, 1 - v)$$

### 2.3. Densidades de cópulas

A pesar de que las cópulas son útiles en la teoría, el realizar un análisis gráfico con ellas no es sencillo. Es por lo anterior que se utilizan de manera frecuente las densidades para ilustrar las distribuciones, más que las gráficas de las funciones de distribución. Cuando una cópula cumple con ser diferenciable, entonces su densidad se encuentra de la siguiente manera, aunque es importante aclarar que no todas las cópulas tienen densidades:

$$c(u_1, u_2) := \frac{\partial^2 C(u_1, u_2)}{\partial u_1 \partial u_2}$$

Si tenemos la cópula expresada de la forma  $C(u_1, u_2) = F(F_1^{\leftarrow}(u_1), F_2^{\leftarrow}(u_2))$ , obtenemos la densidad de cópula en términos de la densidad conjunta junto con las marginales. Debido a que es necesario que la función de distribución sea diferenciable, tenemos que  $F_i^{\leftarrow} = F_i^{-1}$ . Denotando la densidad conjunta por  $f$  y las densidades marginales como  $f_1, f_2$ , tenemos por la regla de la cadena que:

$$c(u_1, u_2) = \frac{f(F_1^{-1}(u_1), F_2^{-1}(u_2))}{f_1(F_1^{-1}(u_1))f_2(F_2^{-1}(u_2))}$$

### 2.4. Cotas para cópulas

Hoeffding y Fréchet, de manera independiente, llegaron a la conclusión de que la cópula siempre recae entre ciertas cotas. La razón de esto, es la existencia de ciertos casos extremos de dependencia.

A continuación, enunciaremos el Teorema de Fréchet-Hoeffding para cópulas:

**Teorema 4.** (*Cotas de Fréchet-Hoeffding*) *Considere una cópula  $C(\mathbf{u}) = C(u_1, u_2)$ . Entonces,*

$$\max\{u_1 + u_2 - 1, 0\} \leq C(\mathbf{u}) \leq \min\{u_1, u_2\}$$

*Demostración.* Sea  $(u_1, u_2) \in I = [0, 1]$ . Para la segunda desigualdad, note que para toda  $i$ :

$$\{U_1 \leq u_1, U_2 \leq u_2\} \subset \{U_i \leq u_i\}$$

Es decir, como:

$$\{U_1 \leq u_1, U_2 \leq u_2\} \subset \{U_1 \leq u_1\} \implies C(u_1, u_2) \leq C(u_1, 1) = u_1$$

y

$$\{U_1 \leq u_1, U_2 \leq u_2\} \subset \{U_2 \leq u_2\} \implies C(u_1, u_2) \leq C(1, u_2) = u_2$$

Por lo que:  $C(u_1, u_2) \leq \min(u_1, u_2)$ .

Para la primera desigualdad, note que:

$$\begin{aligned} C(u_1, u_2) &= \mathbb{P}(\{U_1 \leq u_1\} \cap \{U_2 \leq u_2\}) \\ &= 1 - \mathbb{P}(\{U_1 > u_1\} \cup \{U_2 > u_2\}) \\ &\geq 1 - \sum_{i=1}^2 \mathbb{P}(U_i > u_i) \\ &= 1 - (1 - u_1 + 1 - u_2) \\ &= u_1 + u_2 - 1 \end{aligned}$$

Por lo que:  $C(u_1, u_2) \geq \max(u_1 + u_2 - 1, 0)$ .

□

De hecho, las cotas superior e inferior presentadas en el teorema anterior son cópulas para  $(u_1, u_2) \in [0, 1] \times [0, 1]$ . Se les denota comúnmente como  $M(u_1, u_2) = \min(u_1, u_2)$  y  $W(u_1, u_2) = \max(u_1 + u_2 - 1, 0)$

*Demostración.* Por su definición,  $W$  y  $M$  cumplen con el dominio adecuado. Son ancladas y tienen marginales uniformes, sean  $u_1 \geq 0, u_2 \geq 0$ :

$$\begin{aligned} M(u_1, 0) &= \min(u_1, 0) = 0 \text{ y } M(0, u_2) = \min(0, u_2) = 0 \\ M(u_1, 1) &= \min(u_1, 1) = u_1 \text{ y } M(1, u_2) = \min(1, u_2) = u_2 \end{aligned}$$

Análogamente para  $W$ , sean  $u_1 \leq 1, u_2 \leq 1$ :

$$\begin{aligned} W(u_1, 0) &= \max(u_1 - 1, 0) = 0 \text{ y } W(0, u_2) = \max(u_2 - 1, 0) = 0 \\ W(u_1, 1) &= \max(u_1, 0) = u_1 \text{ y } W(1, u_2) = \max(u_2, 0) = u_2 \end{aligned}$$

Para demostrar que son 2-creciente, sean  $u_1 \leq u_2$  y  $v_1 \leq v_2$ . Para  $M$  tenemos:

$$M(u_2, v_2) - M(u_1, v_2) - M(u_2, v_1) + M(u_1, v_1) = \begin{cases} u_2 - u_1, & \text{si } v_1 \leq u_1 \leq u_2 \leq v_2, \\ v_2 - u_1, & \text{si } v_1 \leq u_1 \leq v_2 \leq u_2, \\ v_2 - v_1, & \text{si } u_1 \leq v_1 \leq v_2 \leq u_2, \\ u_2 - v_1, & \text{si } u_1 \leq v_1 \leq u_2 \leq v_2 \end{cases}$$

Por lo anterior,  $M$  es 2-creciente.

Para ver que  $W$  es 2-creciente, note que  $W(u, v) = 0$  en todo punto  $(u, v)$  tal que  $u + v - 1 \leq 0$  es decir  $u \leq 1 - v$ , por lo que  $W$  se anula bajo  $u = 1 - v$ . Si  $u_1, u_2, v_1, v_2$  están en esta región, tenemos:

$$W(u_2, v_2) - W(u_1, v_2) - W(u_2, v_1) + W(u_1, v_1) = 0$$

Cuando los puntos están en el área delimitada por  $(1, 1)$ ,  $(1, 0)$  y  $(0, 1)$ , entonces:

$$\begin{aligned} & W(u_2, v_2) - W(u_1, v_2) - W(u_2, v_1) + W(u_1, v_1) = \\ & (u_2 + v_2 - 1) - (u_1 + v_2 - 1) - (u_2 + v_1 - 1) + (u_1 + v_1 - 1) = 0 \end{aligned}$$

Si uno de los vértices está en el triángulo superior, tenemos:

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

Pero si dos están, tenemos:

$$\begin{cases} M(u_2, v_2) - M(u_1, v_2) - M(u_2, v_1) + M(u_1, v_1) = \\ \left\{ \begin{array}{ll} u_2 - u_1, & \text{si } (u_2, v_2) \text{ y } (u_1, v_2) \text{ son los vértices que están en la región,} \\ v_2 - u_1, & \text{si } (u_2, v_2) \text{ y } (u_2, v_1) \text{ son los vértices que están en la región} \end{array} \right. \end{cases}$$

Por lo que,  $W$  es 2-creciente. □

## 2.5. Cópulas fundamentales

Las cópulas fundamentales corresponden a tres casos extremos de dependencia entre variables aleatorias: la independencia, la dependencia positiva perfecta y la dependencia negativa perfecta.

La cópula de independencia:

$$\Pi(u_1, u_2) = u_1 \cdot u_2; (u_1, u_2) \in [0, 1] \times [0, 1]$$

relacionada al caso de no dependencia.

**Proposición 4.** *La función definida como:*

$$\Pi(u_1, u_2) = u_1 \cdot u_2; (u_1, u_2) \in [0, 1] \times [0, 1]$$

*es una cópula. Además dos variables aleatorias  $X_1, X_2$  son independientes sí y sólo sí su cópula es la cópula de independencia.*

*Demostración.* Primero demostraremos que  $C$  es una cópula. El  $Dom C$  es por definición  $[0, 1] \times [0, 1]$ , por lo que no hay nada que probar en este aspecto.

$C$  es 2-creciente. Sean  $u_1 \leq u_2$  y  $v_1 \leq v_2$ , entonces:

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

Sus marginales son uniformes ya que  $\Pi(u_1, 1) = u_1$  y  $\Pi(1, u_2) = u_2$ . Cumple con ser anclada ya que  $\Pi(0, u_2) = 0 = \Pi(u_1, 0)$  para toda  $(u_1, u_2)$ . Por lo anterior, podemos concluir que  $\Pi(u_1, u_2)$  es una cópula.

Demostremos ahora que dos variables aleatorias  $X_1, X_2$  son independientes sí y sólo sí su cópula es la cópula de independencia. Sean  $X_1, X_2$  dos variables aleatorias continuas. Sabemos que  $X_1$  y  $X_2$  son independientes sí y sólo sí para toda  $(x_1, x_2)$ ,  $F(x_1, x_2) = F_1(x_1) \times F_2(x_2)$ .

Ahora, utilizando el teorema de Sklar, tenemos que:

$$C(F_1(x_1), F_2(x_2)) = F_1(x_1) \times F_2(x_2)$$

Realizando la transformación  $u_1 = F_1(x_1)$  y  $u_2 = F_2(x_2)$ , y utilizando la continuidad de  $F_1, F_2$ , la igualdad anterior es equivalente a:  $C(u_1, u_2) = u_1 \times u_2$   $\square$

Como ya se mencionó, las cotas de Fréchet-Hoeffding son también cópulas, donde la cópula de *co-monotonicidad* o la cota superior de Fréchet-Hoeffding está dada por:

$$M(u_1, u_2) = \min\{u_1, u_2\}$$

Co-monotonicidad se refiere al caso de dependencia positiva perfecta. Considere una transformación estrictamente creciente  $T_2$  y sea  $X_2 = T_2(X_1)$ . Usando el Teorema 2 nos daremos cuenta de que estas variables efectivamente tienen la cópula co-monotónica o en otras palabras, sean las variables aleatorias  $X_1, X_2$  y  $F_1, F_2$  sus respectivas funciones de distribución marginal. Sea  $F$  su función de distribución conjunta, si  $X_1 = X_2$  o las variables tienen dependencia positiva perfecta, la cópula correspondiente es:

$$\begin{aligned} F(x_1, x_2) &= \mathbb{P}(X_1 \leq x_1, X_2 \leq x_2) \\ &= \mathbb{P}(X_1 \leq x_1, X_1 \leq x_2) \\ &= \mathbb{P}(X_1 \leq (x_1, x_2)) \end{aligned}$$

Lo que implica que:

$$\min(\mathbb{P}(X_1 \leq x_1), \mathbb{P}(X_1 \leq x_2)) = C(F_1(x_1), F_2(x_2))$$

Esto es  $C(u_1, u_2) = M(u_1, u_2)$ .

El otro extremo está dado por la *contra-monotonicidad*, y la cópula asociada sólo se puede obtener para el caso bidimensional (véase Observación 1). La cópula contra-monotónica o cota inferior de Fréchet-Hoeffding está dada por:

$$W(u_1, u_2) = \max\{u_1 + u_2 - 1, 0\}$$

Este caso hace referencia a la dependencia negativa perfecta en el sentido de que  $X_2 = T(X_1)$  con  $T$  una función estrictamente decreciente. Si  $X_2 = -X_1$  o las variables tienen dependencia negativa perfecta, la función de distribución conjunta:

$$\mathbb{P}(X_1 \leq x_1, X_2 \leq x_2) = \mathbb{P}(X_1 \leq x_1, -X_1 \leq x_2) = \mathbb{P}(X_1 \leq x_1, X_1 > -x_2) = \mathbb{P}(-x_2 \leq X_1 \leq x_1)$$

$$\mathbb{P}(-x_2 \leq X_1 \leq x_1) = \begin{cases} F_{X_1}(x_1) - F_{X_1}(-x_2), & \text{si } -x_2 \leq x_1, \\ 0, & \text{c.o.c.} \end{cases}$$

Con lo que podemos concluir que:

$$\mathbb{P}(-x_2 \leq X_1 \leq x_1) = \text{máx} \{F_{X_1}(x_1) - F_{X_1}(-x_2), 0\}$$

La marginal de  $-X_1$ , así como su inversa están dadas por:

$$F_{X_2}(x_2) = F_{-X_1}(x_2) = \mathbb{P}(-X_1 \leq x_2) = \mathbb{P}(X_1 > -x_2) = 1 - \mathbb{P}(X_1 \leq -x_2) = 1 - F_{X_1}(-x_2)$$

y

$$F_{X_2}^{-1}(x_2) = F_{-X_1}^{-1}(u_2) = -F_{X_1}^{-1}(-u_2 + 1)$$

Finalmente,

$$\begin{aligned} F_{X_1, X_2}(F_{X_1}^{-1}(u_1), F_{-X_1}^{-1}(u_2)) &= \text{máx} \left\{ F_{X_1}(F_{X_1}^{-1}(u_1)) - F_{X_1}(-F_{-X_1}^{-1}(u_2)), 0 \right\} \\ &= \text{máx} \left\{ u_1 - F_{X_1}(-(-F_{X_1}^{-1}(-u_2 + 1))), 0 \right\} \\ &= \text{máx} \{u_1 + u_2 - 1, 0\} \end{aligned}$$

Para entender mejor estos conceptos veamos el siguiente ejemplo. Considere dos variables aleatorias uniformes  $U_1, U_2$ . Ciertamente, si  $U_1 = U_2$  las variables son extremadamente dependientes una de la otra. En este caso, la cópula está dada por:

$$\begin{aligned} C(u, v) &= \mathbb{P}(U_1 \leq u_1, U_2 \leq u_2) \\ &= \mathbb{P}(U_1 \leq u_1, U_1 \leq u_2) \\ &= \text{mín}(u_1, u_2) \end{aligned}$$

Es claro que la independencia es completamente contraria a la co-monotonicidad. Sin embargo, la independencia sólo es un paso para el caso contrario extremo de la co-monotonicidad, llamado contra-monotonicidad. En términos de variables aleatorias uniformes, este caso se obtiene cuando  $U_2 = 1 - U_1$ . La cópula relacionada para  $1 - u_2 < u_1$  es:

$$C(u_1, u_2) = \mathbb{P}(U_1 \leq u_1, 1 - U_1 \leq u_2) = \mathbb{P}(U_1 \leq u_1, 1 - u_2 \leq U_1) = u_1 + u_2 - 1$$

y 0 en cualquier otro caso.

**Observación 1.** *Es importante mencionar que la cópula co-monotónica existe en cualquier dimensión  $d$  pero no hay cópula contra-monotónica en el caso de dimensiones mayores a 2.*

*Para explicar esto, considere las variables aleatorias  $X_1, X_2, X_3$ . Somos libres de elegir  $X_1$  y  $X_2$  como contra-monotónicas así como  $X_1$  y  $X_3$ . Sin embargo, esto ya establece una restricción sobre la relación entre  $X_2$  y  $X_3$ . En particular, si  $X_1$  decrece, ambas  $X_2$  y  $X_3$  tienen que crecer, así que ya no podrán ser contra-monotónicas nuevamente. Por otro lado, aún cuando la cópula contra-monotónica no existe, la desigualdad se mantiene.*

En el caso en el que las variables aleatorias tengan distribuciones marginales continuas, tenemos un resultado que nos ayuda a caracterizar de manera simple y potente a la co-monotonicidad:

**Corolario 2.** *Sean  $X_1, X_2$  variables aleatorias con funciones de distribución continuas. Decimos que las variables son co-monotónicas si y sólo  $X_2 = T(X_1)$  casi seguramente para alguna transformación creciente  $T$ .*

En el caso en que  $X_1$  y  $X_2$  son continuas, la contra-monotonicidad es equivalente a que  $X_2 = T(X_1)$  casi seguramente para alguna función decreciente  $T$ .

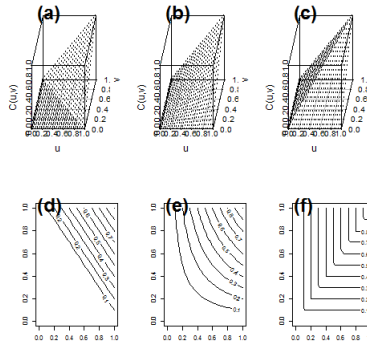


Figura 1: Gráficas de perspectiva (a)–(c) y nivel de las cópulas fundamentales (d)–(f): (a), (d) contra-monotonicidad, (b), (e) independencia y (c), (f) co-monotonicidad.

## 2.6. Cópulas de supervivencia

La probabilidad de que un individuo viva o sobreviva más allá del tiempo  $x$  está dada por la función de supervivencia (o función de confiabilidad)  $\bar{F}(x) = \mathbb{P}(X > x) = 1 - F(x)$ , donde  $F$  es la función de distribución de  $X$ . Cuando hablamos de tiempos de vida, el rango natural de una variable aleatoria es  $[0, \infty)$ ; sin embargo usaremos el término función de supervivencia incluso cuando el rango sean los reales extendidos,  $\bar{\mathbb{R}} = [-\infty, \infty]$ .



La función de supervivencia conjunta  $\bar{F}$  está dada por  $\bar{F}(x_1, x_2) = \mathbb{P}(X_1 > x_1, X_2 > x_2)$ . Las marginales de  $\bar{F}$  son las funciones  $\bar{F}(x_1, -\infty)$  y  $\bar{F}(-\infty, x_2)$  que son las funciones de supervivencia univariada  $\bar{F}_1$  y  $\bar{F}_2$ , respectivamente.

Es natural preguntarse si ¿Existe una relación entre las funciones de supervivencia univariadas y las funciones de supervivencia conjuntas, análoga a la relación existente ente las funciones univariadas y las funciones de distribución conjunta como explica el Teorema de Sklar?. Para responder a la pregunta, suponga que la cópula de  $X_1$  y  $X_2$  es  $C$ , luego:

$$\begin{aligned}\bar{F}(x_1, x_2) &= 1 - F_1(x_1) - F_2(x_2) + F(x_1, x_2) \\ &= \bar{F}_1(x_1) + \bar{F}_2 - 1 + C(F_1(x_1), F_2(x_2)) \\ &= \bar{F}_1(x_1) + \bar{F}_2 - 1 + C(1 - \bar{F}_1(x_1), \bar{F}_2(x_2))\end{aligned}$$

Si definimos una función  $\hat{C}$  de  $I^2$  en  $I$ :

$$\hat{C}(u, v) = u + v - 1 + C(1 - u, 1 - v)$$

tenemos,  $\bar{F}(x_1, x_2) = \hat{C}(\bar{F}_1(x_1), \bar{F}_2(x_2))$ . Nos referiremos a  $\hat{C}$  como la cópula de supervivencia de  $X_1$  y  $X_2$ . Note que  $\hat{C}$  junta a la función conjunta de supervivencia con sus marginales de una manera completamente análoga en la que una cópula conecta la función de distribución conjunta con sus marginales.

Por lo anterior, es claro que existe una versión del teorema de Sklar para las funciones de supervivencia.:

**Corolario 3.** *Sea  $X = (X_1, X_2)$  un vector aleatorio con función de supervivencia bivariada asociada  $\bar{F}$ , funciones de distribución marginales  $F_1, F_2$  y funciones marginales de supervivencia  $\bar{F}_1, \bar{F}_2$ , tenemos la siguiente identidad:*

$$\bar{F}(x_1, x_2) = \hat{C}(\bar{F}_1, \bar{F}_2)$$

En el caso de que  $F_1, F_2$  sean continuas la identidad se establece notando que:

$$\bar{F}(x_1, x_2) = \mathbb{P}(X_1 > x_1, X_2 > x_2) = \mathbb{P}(1 - F_1(X_1) \leq \bar{F}_1(x_1), 1 - F_2(X_2) \leq \bar{F}_2(x_2))$$

Así, el teorema de Sklar para cópulas de supervivencia se sigue escribiendo  $\hat{C}$  para la función de distribución de  $U = (F_1(X_1), F_2(X_2))$ . En general, el término cópula de supervivencia de una cópula  $C$  puede ser usada para denotar la distribución de  $1 - U$  cuando  $U$  tiene función de distribución  $C$ .

## 2.7. Simetría

Si  $X$  es una v.a. y  $a$  es un número real, decimos que  $X$  es simétrica respecto a  $a$  si las funciones de distribución de las variables aleatorias  $X - a$  y  $a - X$  son las mismas, esto es, para cualquier  $x$  en  $\mathbb{R}$ ,  $\mathbb{P}(X - a \leq x) = \mathbb{P}(a - X \leq x)$ . Cuando  $X$  es continua con función de distribución  $F$ , lo anterior es equivalente a:

$$F(a + x) = \bar{F}(a - x)$$

Pero, para el caso bivariado, ¿Qué significa que la pareja  $(X_1, X_2)$  sea simétrica en el punto  $(a, b)$ ? Existen demasiadas respuestas a esta interrogante y cada una de ellas conduce a un diferente tipo de simetría bivariada.

**Definición 9.** Sean  $X_1$  y  $X_2$  variables aleatorias y sea  $(a, b)$  sea un punto en  $\mathbb{R}^2$ .

1.  $(X_1, X_2)$  es marginalmente simétrica alrededor de  $(a, b)$  si  $X_1$  y  $X_2$  son simétricas alrededor de  $a$  y  $b$ , respectivamente.
2.  $(X_1, X_2)$  es radialmente simétrica alrededor de  $(a, b)$  si la función de distribución conjunta de  $X_1 - a$  y  $X_2 - b$  es la misma que la función de distribución conjunta de  $a - X_1$  y  $b - X_2$
3.  $(X_1, X_2)$  es conjuntamente simétrica alrededor de  $(a, b)$  si las siguientes cuatro parejas de variables aleatorias tienen una distribución conjunta común:  $(X_1 - a, X_2 - b)$ ,  $(X_1 - a, b - X_2)$ ,  $(a - X_1, X_2 - b)$  y  $(a - X_1, b - X_2)$ .

Cuando  $X_1, X_2$  son continuas podemos escribir la condición para la simetría radial en términos de la función de distribución conjunta y la de supervivencia:

**Teorema 5.** Sean  $X_1$  y  $X_2$  variables aleatorias continuas con función de distribución conjunta  $F$  y funciones de distribución marginales  $F_1$  y  $F_2$ , respectivamente. Sea  $(a, b)$  sea un punto en  $\mathbb{R}^2$ . Entonces,  $(X_1, X_2)$  es radialmente simétrico alrededor de  $(a, b)$  si y sólo si:  $F(a+x, b+y) = \hat{F}(a-x, b-y)$  para toda  $(x, y)$  en  $\mathbb{R}^2$ .

Se puede probar que la simetría conjunta implica la simetría radial y ésta a su vez, la marginal. En esta ocasión, nos enfocaremos en la simetría radial. Es natural preguntarse si las cópulas y las cópulas de supervivencia juegan un papel importante en la simetría, la respuesta a esta pregunta se encuentra en el siguiente teorema:

**Teorema 6.** Sean  $X_1$  y  $X_2$  variables aleatorias continuas con función de distribución conjunta  $F$  y funciones de distribución marginales  $F_1$  y  $F_2$ , con cópula  $C$ . Además, suponga que  $X_1$  y  $X_2$  son simétricas alrededor de  $a$  y  $b$ , respectivamente. Entonces  $(X_1, X_2)$  es radialmente simétrico alrededor  $(a, b)$ , si y sólo si  $C = \hat{C}$ , i.e. si y sólo si  $C$  satisface la ecuación funcional:

$$C(u, v) = u + v - 1 + C(1 - u, 1 - v) \text{ para toda } (u, v) \text{ en } I^2$$

*Demostración.* Recuerde antes las siguientes definiciones vistas anteriormente:

- La cópula de supervivencia de  $X, Y$  es:  $\hat{C}(u, v) = u + v - 1 + C(1 - u, 1 - v)$
- $X$  es simétrica alrededor de  $a$  si la función de distribución de las variables  $X - a$  y  $a - X$  es la misma, cuando  $X$  es continua entonces hablamos de  $F(a + x) = \bar{F}(a - x)$

Usando lo anterior se tiene el siguiente desarrollo:

$$\begin{aligned}
F(a + x_1, b + x_2) &= \bar{F}(a - x_1, b - x_2) \text{ para toda } (x_1, x_2) \in \mathbb{R}^2 \\
&\Leftrightarrow C(F_1(a + x_1), F_2(b + x_2)) = \hat{C}(\bar{F}_1(a - x_1), \bar{F}_2(b - x_2)) \text{ para toda } (x_1, x_2) \in \mathbb{R}^2 \\
&\Leftrightarrow C(F_1(a + x_1), F_2(b + x_2)) = \hat{C}(F_1(a - x_1), F_2(b - x_2)) \text{ para toda } (x_1, x_2) \in \mathbb{R}^2 \\
&\Leftrightarrow C(u, v) = \hat{C}(u, v) \text{ para toda } (u, v) \in I^2
\end{aligned}$$

□

Otro tipo de simetría es la intercambiabilidad. Las variables aleatorias  $X_1, X_2$  son intercambiables si los vectores  $(X_1, X_2)$  y  $(X_2, X_1)$  se distribuyen de manera idéntica, esto es, si la función de distribución conjunta de  $(X_1, X_2)$  es  $F$  entonces  $F(x_1, x_2) = F(x_2, x_1)$  para todo  $x_1, x_2 \in \mathbb{R}^2$ . Las variables aleatorias intercambiables deben ser idénticamente distribuidas, es decir, deben tener una función de distribución común. Para variables aleatorias idénticamente distribuidas, la intercambiabilidad es equivalente a la simetría de su cópula como se expresa en el siguiente teorema:

**Teorema 7.** Sean  $X_1$  y  $X_2$  variables aleatorias continuas con función de distribución conjunta  $F$  y funciones de distribución marginales  $F_1$  y  $F_2$ , con cópula  $C$ . Entonces,  $X_1$  y  $X_2$  son intercambiables si y sólo si  $F_1 = F_2$  y  $C(u, v) = C(v, u)$  para toda  $(u, v)$  en  $I^2$ .

## 2.8. Conclusiones

- Conociendo la función de distribución conjunta podemos conocer las funciones de distribución marginales. Sin embargo, el conocer las funciones de distribución marginales no nos permite construir a la función de distribución conjunta. Aún nos hace falta conocer la manera en la que se inter-relacionan éstas funciones de distribución marginales, la cópula es la función que nos hace falta y que cumple con esta función.
- El Teorema de Sklar demuestra que toda función bivariada tiene una cópula asociada y brinda un método para construir funciones de distribución multivariada con la ayuda de las cópulas y las funciones de distribución marginales.
- Las cópulas permiten mayor flexibilidad para modelar: podemos modelar a las marginales de manera separada a la estructura de dependencia.
- La cópula es invariante bajo transformaciones estrictamente crecientes, a diferencia de la correlación que es invariante sólo bajo transformaciones lineales.

- Hoeffding y Fréchet determinaron que toda cópula recae entre ciertas cotas, debido a la existencia de casos de dependencia extrema. Las cópulas fundamentales son tres casos de dependencia extrema: independencia, dependencia positiva perfecta (co-monotonicidad) y dependencia negativa perfecta (contra-monotonicidad). La cópula co-monotónica existe para toda dimensión  $d \geq 2$  pero la contra-monotónica sólo para el caso bidimensional.

### 3. Familias de cópulas

En la sección *Cópulas fundamentales* ya vimos tres cópulas que representan un número de estructuras de dependencia importantes. A continuación abordaremos algunas familias de cópulas que pueden ser agrupadas en dos grandes categorías: las cópulas implícitas que se extraen de estructuras de dependencia bien conocidas utilizando el teorema de Sklar, pero no necesariamente poseen una forma cerrada y las cópulas explícitas que tienen formas sencillas y cerradas y siguen construcciones matemáticas generales conocidas para producir cópulas.

#### 3.1. Cópulas elípticas

Dentro de las cópulas implícitas, tenemos a las cópulas asociadas a las distribuciones elípticas. Su característica más importante es que representan relaciones de dependencia simétricas, es decir, es igual analizar la cola izquierda o derecha de las distribuciones asociadas.

Algunos ejemplos son la cópula Gaussiana y la t-Cópula.

Este tipo de cópula cuenta con la siguiente forma:

$$\begin{aligned} C_\rho(u, v) &= \frac{1}{\sqrt{1-\rho^2}} \int_{-\infty}^{\Phi_{g_1}^{-1}(u)} \int_{-\infty}^{\Phi_{g_2}^{-1}(v)} g\left(\frac{x_1^2 - 2x_1x_2\rho + x_2^2}{\sqrt{1-\rho^2}}\right) dx_1 dx_2 \\ &= H_\rho(\Phi_{g_1}^{-1}(u), \Phi_{g_2}^{-1}(v)) \end{aligned}$$

Donde  $g$  es la distribución conjunta de variables  $X, Y$ , mientras que  $\Phi_{g_1}^{-1}(u)$  y  $\Phi_{g_2}^{-1}(v)$  son funciones cuantiles de  $g_1$  y  $g_2$ ,  $\rho$  es su coeficiente de correlación.

#### Cópula Gaussiana

Esta cópula representa la estructura de dependencia de una distribución normal bivariada, es decir, las distribuciones marginales normales con la cópula gaussiana forman las distribuciones normales bivariadas.

Si se desea conservar la estructura de dependencia gaussiana pero las distribuciones marginales no normales se le llama distribución meta Gaussiana.

La distribución normal multivariada conduce a la cópula Gaussiana. Considere dos variables aleatorias que se distribuyen normal  $X_1, X_2$  y que de manera conjunta también se distribuyen normal. Es bien conocido que su correlación describe de manera completa su estructura de dependencia. Esta propiedad se mantiene como válida en *toda* la familia de distribuciones elípticas. Sin embargo es común que se utilice a la correlación fuera de esta familia y establecer resultados incorrectos.

Siguiendo (2), obtenemos la cópula Gaussiana bi-dimensional:

$$C_\rho^{Ga}(u_1, u_2) = \Phi_\Sigma(\Phi^{-1}(u_1), \Phi^{-1}(u_2)),$$

donde  $\Sigma$  es la matriz de  $2 \times 2$  con 1 en la diagonal y  $\rho$  en otro caso.  $\Phi$  denota la función de distribución de una normal mientras que  $\Phi_\Sigma$  es la función de distribución de una normal bivariada con media cero y matriz de varianzas-covarianzas  $\Sigma$ . Note que la representación anterior es equivalente a:

$$\int_{-\infty}^{\Phi^{-1}(u_1)} \int_{-\infty}^{\Phi^{-1}(u_2)} \frac{1}{2\pi\sqrt{1-\rho^2}} \exp\left(-\frac{s_1^2 - 2\rho s_1 s_2 + s_2^2}{2(1-\rho^2)}\right) ds_1 ds_2$$

Para las distribuciones elípticas, independencia es equivalente a correlación cero. Cuando:

- $\rho = 0$ , la cópula Gaussiana es la cópula de independencia
- $\lim_{\rho \rightarrow 1} C_\rho^{Ga}(u_1, u_2)$ , obtenemos la cópula co-monotónica
- $\lim_{\rho \rightarrow -1} C_\rho^{Ga}(u_1, u_2)$ , obtenemos la cópula contra-monotónica

Cuando una cópula se interpola entre tres estructuras de dependencia por medio de un sólo parámetro, se le llama *comprehensiva*, en este caso por ejemplo, el parámetro es  $\rho$ .

Note que la matriz de covarianza  $\Sigma$  utilizada en el ejemplo, es una matriz obtenida de una matriz de covarianza arbitraria re-escalando cada componente tal que la varianza sea 1. De acuerdo al *Teorema 2* esta transformación no cambia la cópula.

Es decir, partiremos de una matriz  $\hat{\Sigma}$ :

$$\hat{\Sigma} = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}$$

y re-escalando cada componente, llegamos a que:

$$\Sigma = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}$$

La densidad de cópula se obtiene derivando la expresión,

$$c(u_1, u_2) = \frac{1}{\sqrt{1 - \rho^2}} \exp\left(-\frac{\rho^2 \zeta_1^2 - 2\rho \zeta_1 \zeta_2 + \rho^2 \zeta_2^2}{2(1 - \rho^2)}\right)$$

donde  $\zeta_1 = \phi^{-1}(u_1)$ ,  $\zeta_2 = \phi^{-1}(u_2)$  son los cuantiles de variables normales  $N(0, 1)$ .

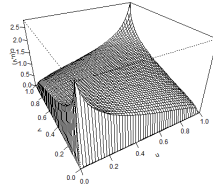


Figura 2: Densidad de la cópula Gaussiana bivariada con correlación 0.3.

### Cópula t de student

La distribución t-Student es llamada una mezcla de distribuciones normales. Recordemos que una variable aleatoria  $\eta$  de una t-distribución con  $\nu$  grados de libertad siempre se puede representar como :

$$\eta = \frac{X_1}{\sqrt{\frac{\xi}{\nu}}} = \frac{\sqrt{\nu} X_1}{\sqrt{Y_1^2 + \dots + Y_\nu^2}}$$

donde  $X_1, \mathbf{Y} = Y_1, \dots, Y_\nu$  son normal estándar, mientras que  $\xi$  tiene distribución  $\chi_\nu^2$ ,  $X_1$  independiente de  $\mathbf{Y}$ . La t-distribución en 2 dimensiones con  $\nu$  grados de libertad se obtiene de  $\mathbf{X} = X_1, X_2 \sim N(0, \sigma)$  vía:

$$(\eta_1, \eta_2) = \left( \frac{X_1}{\sqrt{\frac{\xi}{\nu}}}, \frac{X_2}{\sqrt{\frac{\xi}{\nu}}} \right)$$

donde  $\zeta$  es  $\chi_\nu^2$  independiente de  $\mathbf{X}$ . Formalmente, la t-cópula o cópula Student está dada por:  $C_{\nu, \Sigma}^t(\mathbf{u}) = t_{\nu, \Sigma}(t_\nu^{-1}(u_1), \dots, t_\nu^{-1}(u_d))$  donde  $\Sigma$  es una matriz de correlación,  $t_\nu$  es la función de distribución de una t de Student unidimensional y  $t_{\nu, \Sigma}$  es la función de distribución de una t de Student bivariada. La matriz de correlación se obtiene de una matriz arbitraria escalando cada componente para que tenga varianza 1, como en el caso de la cópula gaussiana. Si la correlación entre  $X_1$  y  $X_2$  es 1 (o -1) tenemos la cópula de co-monotonicidad (contra-monotonicidad), pero es importante notar que cuando  $\rho = 0$  no tenemos independencia, ya que se introduce dependencia vía  $\zeta$ .

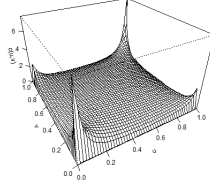


Figura 3: Densidad de la t-cópula bivariada con correlación 0.3 y dos grados de libertad.

### 3.2. Cópulas arquimedianas

Mientras que las cópulas Gaussiana y t son implicadas por distribuciones bien conocidas y no tienen una forma cerrada explícita, podemos encontrar un número de cópulas que tienen una forma cerrada y simple. Este tipo de cópulas permite, a diferencia de la cópula Gaussiana que es simétrica, considerar otros tipos de estructura de dependencia.

**Ejemplo 1.** *Observe los siguientes ejemplos:*

- *La cópula Gumbel bivariada o Gumbel-Hougaard está dada por:*

$$C_{\theta}^{Gu}(u_1, u_2) = \exp[-((- \ln(u_1))^{\theta} + (- \ln(u_2))^{\theta})^{\frac{1}{\theta}}] \text{ con } \theta \in [1, \infty).$$

*Para  $\theta = 1$  tenemos la cópula de independencia, cuando  $\theta \rightarrow \infty$  la cópula tiende a la cópula de co-monotonicidad.*

- *La cópula Clayton está dada por:*

$$C_{\theta}^{Cl}(u_1, u_2) = (\max\{u_1^{-\theta} + u_2^{-\theta} - 1, 0\})^{-\frac{1}{\theta}} \text{ con } \theta \in [-1, \infty) \setminus \{0\}.$$

*Cuando  $\theta \rightarrow 0$  tenemos la cópula de independencia, cuando  $\theta \rightarrow \infty$  la cópula tiende a la cópula de co-monotonicidad, cuando  $\theta = -1$  tenemos la cota inferior de Fréchet-Hoeffding.*

Estos ejemplos sugieren que es posible generar un número de cópulas interpolando entre algunas cópulas de una manera específica. De los dos ejemplos anteriores, uno puede percatarse que la forma general de la cópula es:

$$C(u_1, u_2) = \phi^{-1}(\phi(u_1) + \phi(u_2))$$

donde  $\phi$  es una función decreciente que mapea el  $[0, 1]$  al  $[0, \infty]$ . La función  $\phi$  es llamada **función generadora de la cópula**.

**Teorema 8.** (*Cópula arquimediana*) Considere una función continua y estrictamente decreciente  $\phi : I = [0, 1] \rightarrow [0, \infty)$  con  $\phi(1) = 0$ . Entonces:

$$C(u_1, u_2) = \begin{cases} \phi^{-1}(\phi(u_1) + \phi(u_2)), & \text{si } \phi(u_1) + \phi(u_2) \leq \phi(0), \\ 0 & , \text{ c.o.c} \end{cases}$$

Es una cópula, si y sólo si  $\phi$  es convexa.

Si  $\phi(0) = \infty$  se dice que la generadora es estricta y entonces:

$$C(u_1, u_2) = \phi^{-1}(\phi(u_1) + \phi(u_2)) \quad (3)$$

Bajo este enfoque se puede generar gran número de cópulas, a continuación se muestran algunos ejemplos:

**Definición 10.** Sea  $\phi$  una función continua, estrictamente decreciente que va del  $[0, 1]$  al  $[0, \infty)$  tal que  $\phi(1) = 0$ . La pseudo inversa de  $\phi$  es la función:  $\phi^{[-1]} : [0, \infty) \rightarrow I = [0, 1]$  dada por:

$$\phi^{[-1]}(t) = \begin{cases} \phi^{-1}(t), & \text{si } 0 \leq t \leq \phi(0), \\ 0 & , \text{ si } \phi(0) \leq t \leq \infty \end{cases}$$

Note que  $\phi^{[-1]}$  es continua y no creciente sobre  $[0, \infty]$  y estrictamente decreciente en  $[0, \phi(0))$ . Más aún,

$$\phi(\phi^{[-1]}(t)) = \begin{cases} t & , \text{ si } 0 \leq t \leq \phi(0), \\ \phi(0), & \text{ si } \phi(0) \leq t < \infty \end{cases}$$

Finalmente, si  $\phi(0) = \infty$ , entonces  $\phi^{[-1]} = \phi^{-1}$

**Teorema 9.** Sea  $\phi$  una función estrictamente creciente y continua de  $I$  a  $[0, \infty]$  tal que  $\phi(1) = 0$  y sea  $\phi^{[-1]}$  la pseudo inversa de  $\phi$ . Sea  $C$  la función de  $I^2$  a  $I$  dada por:

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

Luego,  $C$  es una cópula si y sólo si  $\phi$  es convexa.

Véase prueba completa de estos Teoremas en [5].

Si  $\phi(0) = \infty$  decimos que  $\phi$  es una función generadora estricta. En este caso,  $\phi^{[-1]} = \phi^{-1}$  y  $C(u, v) = \phi^{-1}(\phi(u) + \phi(v))$  se dice que es una cópula arquimediana estricta.

**Ejemplo 2.** Sea  $\phi(t) = 1 - t$  para  $t \in I = [0, 1]$ . Luego  $\phi^{[-1]}(t) = 1 - t$  para  $t \in I = [0, 1]$ , y  $0$  para  $t > 1$ ; i.e.  $\phi^{[-1]}(t) = \max(1 - t, 0)$ . Luego  $C(u, v) = \max(u + v - 1, 0) := W(u, v)$ , de donde vemos que la cota inferior bivariada de Fréchet Hoeffding es arquimediana.



Propiedades

**Teorema 10.** *Sea  $C$  una cópula arquimediana con generadora  $\phi$ . Luego:*

1.  $C$  es simétrica, i.e.  $C(u, v) = C(v, u) \forall u, v \in I = [0, 1]$
2.  $C$  es asociativa, i.e.  $C(C(u, v), w) = C(u, C(v, w)) \forall u, v, w \in I = [0, 1]$

A continuación enunciaremos algunos ejemplos de cópulas arquimedianas:

### Cópula de Frank

Sea  $0 \leq \theta \leq \infty$ . La cópula de Frank está dada por:

$$C_\theta(u_1, u_2) = -\frac{1}{\theta} \ln \left\{ 1 + \frac{(e^{-\theta u_1} - 1)(e^{-\theta u_2} - 1)}{e^{-\theta} - 1} \right\}$$

y la función de densidad:

$$c(u_1, u_2) = -\theta g_1 \left( \frac{1 + g_{u+v}}{g_u + g_v + g_1^2} \right)$$

La función generadora es  $\phi(t) = -\ln \frac{e^{-\theta t} - 1}{e^{-\theta} - 1}$  donde  $\theta \in \mathbb{R} \setminus \{0\}$ . Es una cópula estrictamente arquimediana.

### Cópula de Gumbel

La cópula Gumbel está dada por:

$$C_\theta^{Gu}(u_1, u_2) = \exp[-((-\ln(u_1))^\theta + (-\ln(u_2))^\theta)^{\frac{1}{\theta}}]$$

con  $1 \leq \theta < \infty$ .

y la densidad está dada por:

$$c(u_1, u_2; \delta) = (A + \delta - 1)A^{1-2\delta} e^{-A} (u_1 u_2)^{-1} (-\ln(u_1))^{\delta-1} (-\ln(u_2))^{\delta-1}$$

con  $A = [(-\ln(u_1))^\delta + (-\ln(u_2))^\delta]^{\frac{1}{\delta}}$

La función generadora es  $\phi(u) = -(\ln(u_1))^\delta$ ,  $\delta \geq 1$  con inversa  $\phi^{-1}(u) = e^{(-x)^{\frac{1}{\delta}}}$ .

### Cópula de Clayton

Sea  $\theta \in [-1, \infty) \setminus \{0\}$ . La cópula de Clayton se define como:

$$C_\theta^{Cl}(u_1, u_2) = (\max\{u_1^{-\theta} + u_2^{-\theta} - 1, 0\})^{-\frac{1}{\theta}}$$

y la densidad está dada por:

$$c_\theta(u_1, u_2) = u_1^{-(\theta+1)}(1 + \theta)u_2^{-(\theta+1)} \left\{ u_1^{-\theta} + (u_2^{-\theta} - 1) \right\}^{(\theta^{-1}+2)}$$

La función generadora es:

$$\phi(u) = \alpha^{-1}(u^{-\alpha} - 1), \alpha \in [-1, \infty) \setminus \{0\}$$

A continuación veremos las gráficas de las densidades de las cópulas antes presentadas:

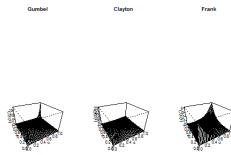


Figura 4: Densidades de las cópulas: Izquierda-Gumbel, Centro-Clayton, Derecha-Frank.

Se debe notar que estas cópulas tienen comportamiento diferente en la esquina izquierda baja y superior derecha, es decir, en los puntos  $(0,0)$  y  $(1,1)$ . La cópula de Gumbel muestra un pico levantado en el  $(1,1)$  y tiene uno menos pronunciado en el  $(0,0)$ . Lo anterior se debe a que la cópula de Gumbel tiene dependencia de cola superior. Para la copula de Clayton la situación es contraria, mientras que la cópula de Frank no se muestra ninguna dependencia de cola.

### 3.3. Cópula bivariada de Marshall Olkin

Introduciremos a esta cópula con un ejemplo práctico. Considere un sistema de dos componentes que están sujetos a descargas, que descomponen a uno u otro componente, ó a ambos. Sean  $X_1, X_2$  los tiempos de vida de cada uno de los componentes. Suponga que las descargas siguen 3 procesos Poisson independientes con parámetros  $\lambda_1, \lambda_2, \lambda_{12} \geq 0$ , donde los índices hacen referencia a si los efectos de las descargas afectan sólo al componente  $i$ , o a ambos.

Los tiempos  $Z_1, Z_2$  y  $Z_{12}$  de ocurrencia de estas descargas son v.a. exponenciales independientes con parámetros  $\lambda_1, \lambda_2, \lambda_{12}$ , respectivamente. Así:

$$\begin{aligned} \bar{H}(x_1, x_2) &= \mathbb{P}(X_1 > x_1, X_2 > x_2) \\ &= \mathbb{P}(Z_1 > z_1)\mathbb{P}(Z_2 > z_2)\mathbb{P}(Z_{12}) > \max(x_1, x_2)) \end{aligned}$$

1

Las funciones de supervivencia univariadas para  $X_1, X_2$  son:

$$\bar{F}_1(x_1) = e^{-(\lambda_1 + \lambda_{12})x_1}$$

y

$$\bar{F}_2(x_2) = e^{-(\lambda_2 + \lambda_{12})x_2}$$

Más aún, como  $máx(x_1, x_2) = x_1 + x_2 - mín(x_1, x_2)$

$$\begin{aligned}\bar{F}_1(x_1) &= e^{-(\lambda_1 + \lambda_{12})x_1 - (\lambda_2 + \lambda_{12})x_2 + \lambda_{12}mín(x_1, x_2)} \\ &= \bar{F}_1(x_1)\bar{F}_2(x_2)mín(e^{\lambda_{12}x_1}, e^{\lambda_{12}x_2})\end{aligned}$$

Sean:

$$\alpha_1 = \frac{\alpha_{12}}{\alpha_1 + \alpha_{12}}$$

y

$$\alpha_2 = \frac{\alpha_{12}}{\alpha_2 + \alpha_{12}}$$

Luego  $e^{(\lambda_{12}x_1)} = \bar{F}_1(x_1)^{-\alpha_1}$  y  $e^{(\lambda_{12}x_2)} = \bar{F}_2(x_1)^{-\alpha_2}$  y así la cópula de supervivencia de  $(X_1, X_2)$  está dada por:

$$\begin{aligned}\hat{C}(u_1, u_2) &= u_1 * u_2 * mín(u_1^{-\alpha_2}, u_2^{-\alpha_2}) \\ &= mín(u_1^{1-\alpha_1}u_2, u_2^{1-\alpha_2}u_1)\end{aligned}$$

Esta construcción lleva a la siguiente familia:

$$\begin{aligned}C_{\alpha_1, \alpha_2}(u_1, u_2) &= mín(u_1^{1-\alpha_1}u_2, u_2^{1-\alpha_2}u_1) \\ &= \begin{cases} u_1^{1-\alpha_1}u_2, & \text{si } u_1^{\alpha_1} \geq u_2^{\alpha_2} \\ u_2^{1-\alpha_2}u_1, & \text{si } u_1^{\alpha_1} \leq u_2^{\alpha_2} \end{cases}\end{aligned}$$

Esta familia es conocida como Marshall-Olkin. Su densidad tiene dos componentes: uno continuo y otro discreto, ya que:

$$\frac{\partial^2}{\partial u_1 \partial u_2} C_{\alpha_1, \alpha_2}(u_1, u_2) = \begin{cases} u_1^{-\alpha_1}, & \text{si } u_1^{\alpha_1} > u_2^{\alpha_2} \\ u_2^{-\alpha_2}, & \text{si } u_1^{\alpha_1} < u_2^{\alpha_2} \end{cases}$$

El punto de masa se encuentra en la curva  $u_1^{\alpha_1} = u_2^{\alpha_2}$  en  $[0, 1]^2$ . Esta cópula tiene una parte absolutamente continua y un componente singular. Para la parte absolutamente continua podemos calcular su densidad y obtenemos:

---

<sup>1</sup> $\mathbb{P}(Z_{12} > máx(x_1, x_2))$  se explica porque el sistema deja de servir cuando ambos componentes fallan

$$\frac{\partial^2}{\partial u_1 \partial u_2} C_{\alpha_1, \alpha_2}(u_1, u_2) = \begin{cases} u_1^{-\alpha_1} & \text{si } u_1^{\alpha_1} > u_2^{\alpha_2} \\ u_2^{-\alpha_2} & \text{si } u_1^{\alpha_1} < u_2^{\alpha_2} \end{cases}$$

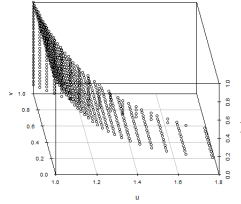


Figura 5: Densidad de la cópula Marshall-Olkin.

### 3.4. Conclusiones

- Analizamos dos clasificaciones de cópulas: implícitas y explícitas. Las primeras se pueden extraer estructuras de dependencia bien conocidas. Entre ellas, incluyen pero no se limitan a las cópulas elípticas, asociadas a las distribuciones elípticas. Representan relaciones de dependencia simétrica sin importar qué cola de las distribuciones asociadas se estudie.
- Las cópulas explícitas, por el contrario, tienen formas sencillas y cerradas. En esta categoría se encuentran las cópulas arquimedianas (se crean utilizando la función generadora de cópulas), por ejemplo: las cópulas Clayton, Gumbel, Frank, entre otras. En este grupo también se encuentra la cópula de Marshall Olkin.

## 4. Medidas de dependencia

En esta sección se revisarán tres tipos de medidas de dependencia: la correlación lineal de Pearson y la correlación de rangos:  $\rho$  de Spearman y  $\tau$  de Kendall. Todas estas medidas producen una medición escalar para la pareja de variables aleatorias  $(X_1, X_2)$  aunque la naturaleza y las propiedades de cada una de las medidas son diferentes en cada caso.

El concepto de correlación es muy importante en la teoría financiera pero vale la pena señalar que solo en el caso de distribuciones elípticas. Como hemos visto, las distribuciones elípticas se describen de manera íntegra por un vector de medias, una matriz de covarianza y una función característica generadora. Debido a que las medias y varianzas son características de las distribuciones marginales, las cópulas de las distribuciones marginales pueden ser pensadas como si dependieran solo de la matriz de correlación y la generadora característica.

Los otros dos tipos de dependencia son medidas basadas en cópulas. En contraste con la correlación ordinaria, estas medidas son funciones de la cópula únicamente y pueden ser usadas en la parametrización de la cópula, como veremos más adelante.

#### 4.1. Correlación lineal

La correlación  $\rho$  entre dos variables  $X_1, X_2$  se define como:

$$\rho(X_1, X_2) = \frac{\text{cov}(X_1, X_2)}{\sqrt{\text{var}(X_1)\text{var}(X_2)}}$$

Es una medida de dependencia lineal y toma valores en  $[-1, 1]$ . Si  $X_1$  y  $X_2$  son independientes, entonces  $\rho(X_1, X_2) = 0$ , pero el opuesto de la afirmación anterior no es cierto.

Si  $|\rho(X_1, X_2)| = 1$ , entonces es equivalente a decir que  $X_2, X_1$  tienen dependencia lineal perfecta, lo que significa que  $X_2 = \alpha + \beta X_1$  casi seguramente para alguna  $\alpha \in \mathbb{R}$  y  $\beta \neq 0$  con  $\beta > 0$  para dependencia lineal positiva y  $\beta < 0$  para dependencia lineal negativa.

Más aún para  $\beta_1, \beta_2 > 0$ ,  $\rho(\alpha_1 + \beta_1 X_1, \alpha_2 + \beta_2 X_2) = \rho(X_1, X_2)$  es decir la correlación es invariante bajo transformaciones lineales estrictamente crecientes. Sin embargo, no es invariante bajo transformaciones estrictamente crecientes no lineales.

Un aspecto que resulta importante remarcar es que la correlación sólo está definida cuando las varianzas de  $X_1$  y  $X_2$  son finitas. Esta restricción a los modelos de varianza finita no es ideal para una medida de dependencia y puede causar problemas cuando se trabaja con distribuciones de cola pesada.

Las siguientes son aseveraciones que son ciertas cuando se trabaja con funciones de distribución elípticas pero que en general no lo son:

- Las distribuciones marginales y las correlaciones dos a dos de un vector aleatorio determinan su distribución conjunta.
- Para distribuciones  $F_1, F_2$  y cualquier valor de correlación  $\rho \in [-1, 1]$  siempre es posible construir la función de distribución conjunta  $F$  con marginales  $F_1$  y  $F_2$  y correlación  $\rho$ .

Si se consideran las afirmaciones anteriores fuera de esta familia se listan algunos inconvenientes:

- En la distribución normal, que la correlación sea 0 es equivalente a independencia, lo que no sucede con la distribución t-student.
- La correlación lineal es invariante bajo transformaciones lineales pero no bajo transformaciones generales. Por ejemplo, dos variables aleatorias log normales tienen una correlación diferente que la de las variables normales subyacentes.

- Para cualquier distribución marginal y dada una correlación  $\rho$ , en general, no es posible re-construir una distribución conjunta (No la original, pero sí alguna).

## 4.2. Correlación de rangos

La correlación de rangos es una medida simple escalar de dependencia que se basa sólo en la cópula de una distribución bivariada y no de las distribuciones marginales, como la correlación lineal que depende de ambas.

Los estimadores empíricos estándar de correlación de rangos se pueden calcular viendo los rangos de los datos, de ahí el nombre. En otras palabras, solo necesitamos conocer el ordenamiento de la muestra para cada variable de interés y no los valores numéricos.

La principal razón práctica para utilizar la correlación de rangos es que pueden ser utilizados para calibrar las cópulas a datos empíricos. Hay dos tipos de correlación de rangos: la de Kendall y la de Spearman.

### Tau de Kendall

La correlación de rangos de Kendall puede ser entendida como una medida de concordancia para vectores aleatorios bivariados. Dos puntos en  $\mathbb{R}^2$ , denotado como  $(x_1, x_2)$  y  $(\bar{x}_1, \bar{x}_2)$ , se dicen concordantes si  $(x_1 - \bar{x}_1)(x_2 - \bar{x}_2) > 0$  y son discordantes si  $(x_1 - \bar{x}_1)(x_2 - \bar{x}_2) < 0$ .

Ahora considere el vector aleatorio  $(X_1, X_2)$  y una copia idéntica independiente  $(\tilde{X}_1, \tilde{X}_2)$  (es decir un vector con la misma distribución pero independiente del primero). Si  $X_2$  tiende a decrecer cuando crece  $X_1$ , entonces esperamos que la probabilidad de concordancia sea muy grande de manera relativa a la probabilidad de discordancia. Si  $X_2$  tiende a decrecer cuando  $X_1$  incrementa, esperamos lo contrario. Esto motiva la correlación de rangos de Kendall, que es simplemente la probabilidad de concordancia menos la probabilidad de discordancia de estas parejas:

$$\rho_\tau(X_1, X_2) = \mathbb{P}((X_1 - \bar{X}_1)(X_2 - \bar{X}_2) > 0) - \mathbb{P}((X_1 - \bar{X}_1)(X_2 - \bar{X}_2) < 0)$$

Esta expresión se puede reducir, lo que nos lleva al siguiente estimador:

**Definición 11.** Para variables aleatorias  $X_1$  y  $X_2$ , la tau de Kendall está dada por:

$$\rho_\tau(X_1, X_2) = E(\text{sign}[(X_1 - \bar{X}_1)(X_2 - \bar{X}_2)])$$

donde  $(\bar{X}_1, \bar{X}_2)$  es una copia independiente de  $(X_1, X_2)$

## Rho de Spearman

Esta medida también puede ser definida en términos de concordancia y discordancia de parejas aleatorias pero la definición más intuitiva para nuestros propósitos es con cópulas.

**Definición 12.** *Para variables aleatorias  $X_1$  y  $X_2$ , la rho de Spearman está dada por:*

$$\rho_S(X_1, X_2) = \rho(F_1(X_1), F_2(X_2))$$

En otras palabras, la rho de Spearman es simplemente la correlación lineal de la probabilidad de variables aleatorias transformadas.

La tau de Kendall y la rho de Spearman tienen diferentes propiedades en común. Ambas son medidas simétricas que toman valores en  $[-1, 1]$ . Dan el valor de cero a las variables independientes, aunque una correlación de rango cero no necesariamente implica independencia. Se puede demostrar que toman el valor de 1 cuando  $X_1$  y  $X_2$  son co-monotónicas y -1 cuando son contra-monotónicas. Ahora demostraremos que, para distribuciones marginales continuas ambas correlaciones de rangos dependen sólo de la única cópula y entonces hereda la propiedad de invarianza bajo transformaciones estrictamente crecientes.

**Proposición 5.** *Suponga que  $X_1$  y  $X_2$  tienen distribuciones marginales continuas y una única cópula  $C$ . Entonces las correlaciones de rangos están dadas por:*

$$\begin{aligned} \rho_\tau(X_1, X_2) &= 4 \int_0^1 \int_0^1 C(u_1, u_2) dC(u_1, u_2) - 1 \\ \rho_\rho(X_1, X_2) &= 12 \int_0^1 \int_0^1 (C(u_1, u_2) - u_1 u_2) du_1 du_2 \end{aligned}$$

*Demostración.* Podemos reescribir la definición de la tau de Kendall como:

$$\rho_\tau(X_1, X_2) = 2\mathbb{P}((X_1 - \bar{X}_1)(X_2 - \bar{X}_2) > 0) - 1$$

y de las propiedades de las parejas  $(X_1, X_2)$  y  $(\bar{X}_1, \bar{X}_2)$ , tenemos:

$$\begin{aligned} \rho_\tau(X_1, X_2) &= 4\mathbb{P}(X_1 < \bar{X}_1, X_2 < \bar{X}_2) - 1 \\ &= 4\mathbb{E}(\mathbb{P}(X_1 < \bar{X}_1, X_2 < \bar{X}_2 | \bar{X}_1, \bar{X}_2)) - 1 \\ &= 4 \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \mathbb{P}(X_1 < x_1, X_2 < x_2) dF(x_1, x_2) - 1 \end{aligned} \tag{4}$$

Como  $X_1$  y  $X_2$  tienen marginales uniformes, inferimos que:

$$\rho_\tau(X_1, X_2) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} C(F_1(x_1), F_2(x_2)) dC(F_1(x_1), F_2(x_2)) - 1 \tag{5}$$

y obtenemos la igualdad deseada sustituyendo  $u_1 = F_1(x_1)$  y  $u_2 = F_2(x_2)$ .

Para la  $\rho$  de Spearman observe que  $\rho_S(X_1, X_2) = 12cov(F_1(X_1), F_2(X_2))$  y como  $F_i(X_i)$  tiene distribución uniforme con varianza  $\frac{1}{12}$ . La segunda igualdad se sigue de aplicar la fórmula de Hoeffding:

$$cov(X_1, X_2) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (F(x_1, x_2) - F_1(x_1)F_2(x_2))dx_1dx_2$$

□

Si regresamos a las aseveraciones falsas presentadas anteriormente<sup>2</sup>, nos damos cuenta de que la primera sigue sin ser cierta. Sin embargo, la segunda sale de la jugada cuando utilizamos correlación de rangos: para cualquier elección de distribuciones marginales continuas es posible especificar una distribución bivariada que tenga la correlación de rangos deseada en  $[-1, 1]$ .

### 4.3. Resumen del capítulo

- Existe una creencia generalizada de que conociendo las marginales y la correlación entre ellas, éstas determinan de manera única a la función de distribución conjunta.
- La correlación lineal tiene algunas debilidades para ser tomada como medida de dependencia, entre ellas se encuentra el hecho de que no es invariante únicamente ante transformaciones lineales (no bajo cualquier tipo de transformación).
- Además, la correlación lineal es importante únicamente en el caso de distribuciones elípticas y está definida únicamente cuando la varianza de ambas variables es finita, lo cual puede representar un obstáculo cuando se trabaja con distribuciones de cola pesada. Es importante mencionar, que en otros casos la correlación podría también ser importante, sin embargo en el caso de distribuciones elípticas el conocer esta medida implica conocer todo sobre la distribución.
- Las tau de Kendall y la rho de Spearman son medidas de asociación que pueden ser expresadas en términos de cópulas, por eso suelen ser una alternativa a la correlación lineal.

## 5. Funciones y coeficientes de dependencia

### 5.1. Coeficientes de dependencia de cola

Son medidas de dependencia dos a dos que dependen únicamente de la cópula de un par de variables aleatorias  $X_1$  y  $X_2$  con funciones de distribución marginales continuas. La motivación de estos coeficientes es que proveen medidas de dependencia extrema, o en otras palabras.

---

<sup>2</sup>Presentadas al final de la sección 4.1.



miden la fuerza de dependencia en las colas de una distribución bivariada. Los coeficientes que describiremos están en términos de los límites de probabilidades condicionales de excedentes de cuantiles.

La dependencia de cola es utilizada para describir qué tan grandes (o pequeños) valores de una variable aleatoria aparecen con grandes (o pequeños) valores en la otra. Este fenómeno es de especial importancia cuando  $X_1$  y  $X_2$  se interpretan como potenciales pérdidas financieras.

El análisis clásico de la dependencia suele incluir sólo la dependencia positiva de cola, sin embargo, en esta ocasión se analizará también la dependencia negativa de cola. La dependencia positiva de cola mide la dependencia entre las variables en el cuadrante superior derecho y en el inferior izquierdo de  $I^2$ , es decir, cuando las variables se encuentran en la misma ubicación del cuadrante.

Con base en lo anterior, es natural preguntarse qué es lo que sucede con la dependencia de las variables cuando se encuentran en los cuadrantes opuestos. A este tipo de dependencia, la llamaremos dependencia negativa de cola, la cual mide la dependencia entre las variables cuando una se encuentra en el cuadrante superior izquierdo y la otra en el cuadrante inferior derecho ó viceversa.

En la dependencia positiva de cola se distingue entre la dependencia positiva de cola superior e inferior. Considere dos variables aleatorias  $U_1, U_2 \sim U(0, 1)$  con cópula  $C$ . La dependencia de cola superior significa de manera intuitiva que a grandes valores de  $U_1$  se esperan también grandes valores de  $U_2$ . De manera más precisa, la probabilidad de que  $U_1$  exceda un límite  $q$  dado que  $U_2$  ha excedido el mismo límite  $q$  cuando  $q \rightarrow 1$  es mayor que cero.

**Definición 13.** Para v.a.'s  $X_1, X_2$  con función de distribución marginal  $F_1, F_2$  definimos el coeficiente de dependencia de cola superior como:

$$\lambda_U = \lim_{q \rightarrow 1} \mathbb{P}(X_2 > F_2^{\leftarrow}(q) | X_1 > F_1^{\leftarrow}(q))$$

siempre y cuando el límite exista y que  $\lambda_U \in I = [0, 1]$ .

El coeficiente de cola inferior se define de manera análoga:

$$\lambda_L = \lim_{q \rightarrow 0} \mathbb{P}(X_2 \leq F_2^{\leftarrow}(q) | X_1 \leq F_1^{\leftarrow}(q))$$

Si  $\lambda_U(\lambda_L) > 0$  entonces hay dependencia de cola. Si  $\lambda_U(\lambda_L) = 0$  decimos que  $X_1, X_2$  son asintóticamente independientes.

Para funciones de distribución continuas, podemos obtener expresiones para los coeficientes de dependencia de cola en términos de la cópula utilizando la regla de Bayes:

$$\lambda_U = \lim_{q \rightarrow 1} \frac{\mathbb{P}(X_2 > F_2^{\leftarrow}(q), X_1 > F_1^{\leftarrow}(q))}{\mathbb{P}(X_1 > F_1^{\leftarrow}(q))}$$

usando que:

$$\mathbb{P}(X_1 > x_1) = \mathbb{P}(X_1 > x_1, X_2 \leq x_2) + \mathbb{P}(X_1 > x_1, X_2 > x_2)$$

y

$$\mathbb{P}(X_2 \leq x_2) = \mathbb{P}(X_1 \leq x_1, X_2 \leq x_2) + \mathbb{P}(X_1 > x_1, X_2 \leq x_2)$$

tenemos que,

$$\mathbb{P}(X_1 > x_1, X_2 > x_2) = 1 - \mathbb{P}(X_1 \leq x_1) - \mathbb{P}(X_2 \leq x_2) + \mathbb{P}(X_1 \leq x_1, X_2 \leq x_2)$$

Además recuerde que

$$\mathbb{P}(X_1 \leq F_1^{\leftarrow}(q)) = F_1(F_1^{\leftarrow}(q)) = q$$

Por lo que nos queda:

$$\begin{aligned} \lambda_U &= \lim_{q \rightarrow 1} \frac{1 - q - q + C(q, q)}{1 - q} \\ &= 2 - \lim_{q \rightarrow 1} \frac{1 - C(q, q)}{1 - q} \end{aligned}$$

o de manera análoga:

$$\lambda_U = 2 - \lim_{q \rightarrow 0} \frac{1 - C(1 - q, 1 - q)}{q}$$

Para el coeficiente de cola inferior, utilizando que  $F_1(X_1) \sim U(0, 1)$  y la definición de cópula obtenemos:

$$\begin{aligned} \lambda_L &= \lim_{q \rightarrow 0} \frac{\mathbb{P}(X_2 \leq F_2^{\leftarrow}(q), X_1 \leq F_1^{\leftarrow}(q))}{\mathbb{P}(X_1 \leq F_1^{\leftarrow}(q))} \\ &= \lim_{q \rightarrow 0} \frac{C(q, q)}{q} \end{aligned}$$

Así, debemos revisar la pendiente de la cópula cuando se acerca al  $(0, 0)$  o al  $(1, 1)$ . Si la pendiente es mayor que 1 (que se refiere al caso de independencia), entonces la cópula exhibe dependencia de cola. Es claro que entre mayor sea la pendiente mayor será la dependencia de cola.

**Ejemplo 3.** *Coefficientes de dependencia de las cópulas Clayton y Gumbel.*

Recordemos que la cópula de Clayton está dada por

$$C_\theta^{Cl} = \left( \max \left\{ u_1^{-\theta} + u_2^{-\theta} - 1, 0 \right\} \right)^{-\frac{1}{\theta}}$$

Por definición, el coeficiente de cola superior de esta cópula es:

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

Utilizando L'Hôpital:

$$\lambda_U = \lim_{q \rightarrow 1} \frac{-2 + 2q^{-\theta-1}(2q^{-\theta} - 1)^{-\theta-1-1}}{-1} = 0$$

Mientras que:

$$\begin{aligned} \lambda_L &= \lim_{q \rightarrow 0} \frac{C(q, q)}{q} \\ &= \lim_{q \rightarrow 0} \frac{(\max\{2q^{-\theta} - 1, 0\})^{\frac{-1}{\theta}}}{q} \\ &= \lim_{q \rightarrow 0} (\max\{2q^{-\theta} - 1, 0\} * q^{\theta})^{\frac{-1}{\theta}} \\ &= 2^{-\frac{1}{\theta}} \end{aligned}$$

En el caso de la cópula de Gumbel, ésta se encuentra dada por la siguiente expresión:

$$C_{\theta}^{Gu}(u_1, u_2) = \exp[-((- \ln(u_1))^{\theta} + (- \ln(u_2))^{\theta})^{\frac{1}{\theta}}]$$

Por lo que sus coeficientes de dependencia son:

$$\begin{aligned} \lambda_U &= 2 - \lim_{q \rightarrow 0} \frac{1 - C(1-q, 1-q)}{q} \\ &= 2 - \lim_{q \rightarrow 0} \frac{1 - \exp[-2^{\frac{1}{\theta}}(-\ln(1-q))]}{q} \\ &= 2 - \lim_{q \rightarrow 0} \frac{(1-q)^{-2^{\frac{1}{\theta}}}}{q} \end{aligned}$$

Aplicando L'Hôpital, queda:

$$\lambda_U = 2 - \lim_{q \rightarrow 0} \frac{(-2^{\frac{1}{\theta}})(1-q)^{-2^{\frac{1}{\theta}}-1}(-1)}{1} = 2 - 2^{\frac{1}{\theta}}$$

Para  $\lambda_L$  tenemos:

$$\lambda_L = \lim_{q \rightarrow 0} \frac{C(q,q)}{q} = \lim_{q \rightarrow 0} \frac{\exp[-(2(-\ln q)^{\theta})^{\frac{1}{\theta}}]}{q} = \lim_{q \rightarrow 0} \frac{[-q]^{-2^{\frac{1}{\theta}}}}{q}$$

Utilizando L'Hôpital, tenemos que:

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

Veamos la siguiente figura:

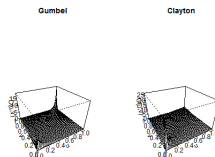


Figura 6: En el lado izquierdo encontramos las gráficas de las densidades de las cópula de Gumbel ( $\theta = 2$ ) y de lado derecho la cópula de Clayton ( $\theta = 2$ )

Se puede notar que las cópulas presentadas tienen comportamientos diferentes en las esquinas superior derecha e inferior izquierda (los puntos  $(1,1)$  y  $(0,0)$ , respectivamente). De manera cualitativa, la copula de Gumbel muestra un pico levantado en  $(1,1)$  mientras que un comportamiento menos pronunciado en  $(0,0)$ . Gracias al análisis de la dependencia de cola anterior, sabemos que la cópula de Gumbel tiene dependencia de cola superior.

Caso contrario es la cópula de Clayton, la cual presenta un pronunciado levantamiento en  $(0,0)$ , lo que refuerza lo obtenido en los cálculos anteriores donde veíamos que esta cópula tiene dependencia de cola inferior.

Debido a que las cópulas Gaussiana y  $t$ , no se encuentran dadas de manera explícita, el cálculo de la dependencia de cola es más complicado. En [4] Sección 5.3.1, se puede encontrar el desarrollo de los cálculos. A continuación se presentan sus resultados:

La cópula Gaussiana es asintóticamente independiente en sus colas superior e inferior. Lo que significa que no importa que correlación escojamos, no tendremos dependencia de cola de una cópula Gaussiana.

Para la  $t$ -Cópula la situación es diferente, el coeficiente de dependencia de cola es:

$$\lambda_U = \lambda_L = 2t_{\nu+1} \left( -\sqrt{\frac{(\nu+1)(1-\rho)}{1+\rho}} \right), \rho > -1$$

Así, la  $t$ -Cópula tiene dependencia de cola superior e inferior, aún cuando  $\rho = 0$ .

La cópula de Gumbel muestra un comportamiento de dependencia de cola superior, mientras que la cópula de Clayton muestra dependencia de cola inferior. La  $t$ -cópula tienen ambos, dependencia de cola inferior y superior. Lo que contrasta con la cópula normal que no exhibe ninguna de las dos dependencias.

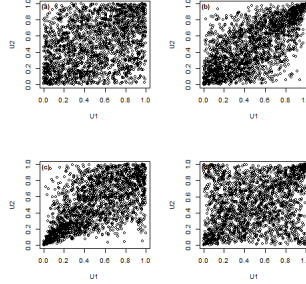


Figura 7: Dos mil simulaciones de las cópulas (a) Gaussiana, (b) Gumbel, (c) Clayton and (d) t

En diversas ocasiones y particularmente al analizar series de tiempo financieras, sucede que las dos series presentan el mismo tipo de comportamiento extremo, ya sea con dependencia superior e inferior. sin embargo, también puede darse el caso de que un valor en la cola izquierda en una serie aparezca al mismo tiempo en el que un valor en la cola derecha aparece en la otra serie. A lo anterior, nos referiremos como dependencia negativa de cola inferior superior (LU) o dependencia negativa de cola superior inferior (UL), por sus siglas en inglés.

**Definición 14.** Sea  $(X_1, X_2)$  un vector aleatorio, sus funciones de probabilidad superior-inferior (UL) e inferior-superior (LU) son  $F_{LU}(x_1, x_2) = \mathbb{P}(X_1 \leq x_1, X_2 > x_2)$  y  $F_{UL}(x_1, x_2) = \mathbb{P}(X_1 > x_1, X_2 \leq x_2)$ . Si  $F_1, F_2, \bar{F}_1, \bar{F}_2$  son las funciones de distribución y supervivencia de  $X_1$  y  $X_2$  respectivamente, nos referiremos a  $F_1$  y  $\bar{F}_2$  como las marginales de  $F_{LU}$  y  $\bar{F}_1$  y  $F_2$  como las marginales de  $F_{UL}$

Las cópulas que consideraremos son  $C_{LU}$  y  $C_{UL}$  que ligan la definición anterior a sus respectivas marginales. Nos referiremos a ellas como las cópulas de las funciones de distribución  $LU$  y  $UL$ , o simplemente como las cópulas  $UL$  y  $LU$ .

Existe una versión del Teorema de Sklar para las cópulas  $UL$  y  $LU$  que garantiza la existencia y unicidad de  $C_{LU}$  y  $C_{UL}$  en el caso continuo.

**Teorema 11.** (Teorema de Sklar para las funciones de distribución  $LU$  y  $UL$ ) Sea  $(X_1, X_2)$  un vector aleatorio,  $F_{LU}$  y  $F_{UL}$  sus funciones de distribución  $LU$  y  $UL$ , respectivamente. Sean  $F_1$  y  $F_2$  las funciones de distribución de  $X_1$  y  $X_2$ , entonces existen cópulas únicas  $C_{LU}$  y  $C_{UL} : [0, 1]^2 \rightarrow [0, 1]$ , tales que, para toda  $X_1, X_2$  en  $[-\infty, \infty]$ :

$$\begin{aligned} F_{LU}(x_1, x_2) &= C_{LU}(F_1(x_1), \bar{F}_2(x_2)) \\ F_{UL}(x_1, x_2) &= C_{UL}(\bar{F}_1(x_1), F_2(x_2)) \end{aligned}$$

Por otro lado, si tenemos cualquier cópula  $C_{LU}$  y  $C_{UL}$ , y  $F_1, F_2, \bar{F}_1, \bar{F}_2$  son las funciones de distribución y supervivencia de  $X_1$  y  $X_2$  respectivamente, entonces considerando las ecuaciones anteriores,  $F_{LU}$  define la función de probabilidad conjunta  $LU$  con marginales  $F_1$  y  $\bar{F}_2$  y  $F_{UL}$  define la función de probabilidad conjunta  $UL$  con marginales  $\bar{F}_1$  y  $F_2$ .

Notemos que el teorema anterior implica que en el caso continuo  $C_{LU}$  y  $C_{UL}$  son las funciones de probabilidad LU y UL de los vectores  $(F_1, \bar{F}_2)$  y  $(\bar{F}_1, F_2)$ , caracterizados como sigue:

$$\begin{aligned} F_{LU}(x_1, x_2) &= C_{LU}(F_1(x_1), \bar{F}_2(x_2)), & C_{LU}(u, v) &= F_{LU}(F_1^{-1}(u), F_2^{-1}(1-v)) \\ F_{UL}(x_1, x_2) &= C_{UL}(\bar{F}_1(x_1), F_2(x_2)), & C_{UL}(u, v) &= F_{UL}(F_1^{-1}(1-u), F_2^{-1}(v)) \end{aligned}$$

Se debe notar que es posible expresar a  $\lambda_L$ ,  $\lambda_U$ ,  $\lambda_L U$  y  $\lambda_U L$ , en términos de la cópula de supervivencia:

$$\begin{aligned} \lambda_L &= \lim_{u \rightarrow 0} \frac{C(u, u)}{u} \\ &= \lim_{p \rightarrow 1} \frac{C(1-p, 1-p)}{1-p} \\ &= \lim_{p \rightarrow 1} \frac{\hat{C}(p, p) - p - p + 1}{1-p} \\ &= \lim_{p \rightarrow 1} \frac{\hat{C}(p, p) - 2p + 1}{1-p} \end{aligned}$$

$$\begin{aligned} \lambda_U &= \lim_{u \rightarrow 1} \frac{1 - 2u + C(u, u)}{1, u} \\ &= \lim_{m \rightarrow 0} \frac{2m - 1 + C(1-m, 1-m)}{m} \\ &= \lim_{m \rightarrow 0} \frac{\hat{C}(m, m)}{m} \end{aligned}$$

## 5.2. Resultados generales

**Proposición 6.** *Sea  $(X_1, X_2)$  un vector aleatorio con funciones de distribución marginales  $F_1, F_2$  y su correspondiente función de distribución  $F$ , así como sus cópulas asociadas  $C, \bar{C}, C_{LU}, C_{UL}$ . Sea  $(U_1, U_2) = (F_1(X_1), F_2(X_2))$ . Luego,*

- $C$  es la función de distribución de  $(U_1, U_2)$
- $\bar{C}$  es la función de distribución de  $(1 - U_1, 1 - U_2)$
- $C_{LU}$  es la función de distribución de  $(U_1, 1 - U_2)$
- $C_{UL}$  es la función de distribución de  $(1 - U_1, U_2)$

*Demostración.* Sean  $U_1, U_2$  variables aleatorias tal que  $U_1, U_2 \sim (0, 1)$

$$\mathbb{P}(U_1 \leq u_1, U_2 \leq u_2) = C(u_1, u_2)$$

■

$$\begin{aligned}
\mathbb{P}(1 - U_1 \leq u_1, 1 - U_2 \leq u_2) &= \mathbb{P}(U_1 > 1 - u_1, U_2 > 1 - u_2) \\
&= 1 - \mathbb{P}(U_1 \leq 1 - u_1) - \mathbb{P}(U_2 \leq 1 - u_2) + \mathbb{P}(U_1 \leq 1 - u_1, U_2 \leq 1 - u_2) \\
&= 1 - (1 - u_1) - (1 - u_2) + C(1 - u_1, 1 - u_2) \\
&= \hat{C}
\end{aligned}$$

■

$$\begin{aligned}
\mathbb{P}(U_1 \leq u_1, 1 - U_2 \leq u_2) &= \mathbb{P}(U_1 \leq u_1, U_2 \geq 1 - u_2) \\
&= \mathbb{P}(U_1 \leq u_1) + \mathbb{P}(U_1 \leq u_1, U_2 \leq 1 - u_2) \\
&= u_1 - C(u_1, 1 - u_2)
\end{aligned}$$

■

$$\begin{aligned}
\mathbb{P}(1 - U_1 \leq u_1, U_2 \leq u_2) &= \mathbb{P}(U_1 \geq 1 - u_1, U_2 \leq u_2) \\
&= \mathbb{P}(U_2 \leq u_2) + \mathbb{P}(U_1 \leq 1 - u_1, U_2 \leq u_2) \\
&= u_2 - C(1 - u_1, u_2)
\end{aligned}$$

□

Para el caso de dependencia negativa, definimos los siguientes coeficientes de dependencia:

$$\begin{aligned}
\lambda_{LU} &= \lim_{q \rightarrow 1} \mathbb{P}(X_2 > q | X_1 < (1 - q)) \\
&= \lim_{q \rightarrow 1} \frac{\mathbb{P}(X_2 > q, X_1 < (1 - q))}{\mathbb{P}X_1 < (1 - q)} \\
&= \lim_{q \rightarrow 1} \frac{\mathbb{P}(X_1 \leq (1 - q)) - \mathbb{P}(X_1 \leq (1 - q), X_2 \leq q)}{\mathbb{P}(X_1 \leq (1 - q))} \\
&= \lim_{q \rightarrow 1} \frac{1 - q - C(1 - q, q)}{1 - q} \\
&= 1 - \lim_{q \rightarrow 1} \frac{C(1 - q, q)}{1 - q}
\end{aligned}$$

De manera análoga:

$$\begin{aligned}
\lambda_{UL} &= \lim_{q \rightarrow 1} \mathbb{P}(X_2 < (1 - q) | X_1 > (q)) \\
&= \lim_{q \rightarrow 1} \frac{\mathbb{P}(X_2 < (1 - q), X_1 > q)}{\mathbb{P}X_1 > q} \\
&= \lim_{q \rightarrow 1} \frac{\mathbb{P}(X_2 \leq (1 - q)) - \mathbb{P}(X_1 \leq q, X_2 \leq (1 - q))}{1 - \mathbb{P}(X_2 \leq q)} \\
&= \lim_{q \rightarrow 1} \frac{1 - q - C(q, 1 - q)}{1 - q} \\
&= 1 - \lim_{q \rightarrow 1} \frac{C(q, 1 - q)}{1 - q}
\end{aligned}$$

Es frecuente encontrar en la literatura que  $\lambda_{UU} = \lambda_U > 0$  represente la existencia de dependencia de cola asintótica superior superior entre los eventos extremos  $\{X_2 > F_2^{-1}(q)\}$  y  $\{X_1 > F_1^{-1}(q)\}$ .

De la misma manera,  $\lambda_{LL} = \lambda_L > 0$  represente la existencia de dependencia de cola asintótica inferior inferior entre los eventos extremos  $\{X_2 < F_2^{-1}(1-q)\}$  y  $\{X_1 < F_1^{-1}(1-q)\}$ .

Haciendo una analogía,  $\lambda_{UL} > 0$  significa que existe dependencia de cola superior inferior asintótica entre los eventos  $\{X_2 < F_2^{-1}(1-q)\}$  y  $\{X_1 > F_1^{-1}(q)\}$ . También  $\lambda_{LU} > 0$  significa que existe dependencia asintótica de cola inferior superior entre los eventos extremos  $\{X_2 > F_2^{-1}(q)\}$  y  $\{X_1 < F_1^{-1}(1-q)\}$ .

Las funciones de dependencia extrema son funciones que describen la estructura de dependencia de distribuciones bivariadas en la cola y se definen a continuación: Sea  $(X_1, X_2)$ , definimos:

$$\begin{aligned}\Lambda_U(x_1, x_2) &:= X_2 \lim_{t \rightarrow \infty} \mathbb{P}(X_1 > F_1^{\leftarrow}(1 - \frac{x_1}{t}) | X_2 > F_2^{\leftarrow}(1 - \frac{x_2}{t})) \\ \Lambda_L(x) &:= X_2 \lim_{t \rightarrow \infty} \mathbb{P}(X_1 \leq F_1^{\leftarrow}(\frac{x_1}{t}) | X_2 \leq F_2^{\leftarrow}(\frac{x_2}{t}))\end{aligned}$$

Note que:

$$\begin{aligned}\Lambda_U(1, 1) &= \lim_{t \rightarrow \infty} \frac{\mathbb{P}(X_1 > F_1^{\leftarrow}(1 - \frac{1}{t}), X_2 > F_2^{\leftarrow}(1 - \frac{1}{t}))}{\mathbb{P}(X_2 > F_2^{\leftarrow}(1 - \frac{1}{t}))} \\ &= \lim_{u \rightarrow 1^-} \frac{\mathbb{P}(X_1 > F_1^{\leftarrow}(u), X_2 > F_2^{\leftarrow}(u))}{\mathbb{P}(X_2 > F_2^{\leftarrow}(u))} \\ &= \lambda_U\end{aligned}$$

$$\begin{aligned}\Lambda_L(1, 1) &= \lim_{t \rightarrow \infty} \frac{\mathbb{P}(X_1 \leq F_1^{\leftarrow}(\frac{1}{t}), X_2 \leq F_2^{\leftarrow}(\frac{1}{t}))}{\mathbb{P}(X_2 \leq F_2^{\leftarrow}(\frac{1}{t}))} \\ &= \lim_{u \rightarrow 0} \frac{\mathbb{P}(X_1 \leq F_1^{\leftarrow}(u), X_2 \leq F_2^{\leftarrow}(u))}{\mathbb{P}(X_2 \leq F_2^{\leftarrow}(u))} \\ &= \lambda_L\end{aligned}$$

La función de dependencia de cola así como el coeficiente de dependencia no dependen de las marginales pero si de la cópula, por ejemplo:  $\lambda_U = \lim_{u \rightarrow 1^-} \frac{1-2u+C(u,v)}{1-u} \Lambda_U(x_1, x_2) = x_1 + x_2 - \lim_{t \rightarrow \infty} t(1 - C(1 - \frac{x_1}{t}, 1 - \frac{x_2}{t}))$

### 5.3. Funciones y coeficientes de dependencia generales

Recordemos que:

$$\lambda_L = \lim_{q \rightarrow 0} \frac{C(q, q)}{q}$$



y también:

$$\begin{aligned}
\lambda_U &= \lim_{q \rightarrow 1} \frac{1 - q - q + C(q, q)}{1 - q} \\
&= \lim_{q \rightarrow 1} \frac{\hat{C}(1 - q, 1 - q)}{1 - q} \\
&= \lim_{m \rightarrow 0} \frac{\hat{C}(m, m)}{m}
\end{aligned}$$

Además:

$$\begin{aligned}
\lambda_{LU} &= 1 - \lim_{q \searrow 1} \frac{C(1 - q, q)}{1 - q} \\
&= \lim_{q \searrow 1} \frac{(1 - q) - C(1 - q, q)}{1 - q} \\
&= \lim_{q \searrow 1} \frac{C_{LU}(1 - q, 1 - q)}{1 - q} \\
&= \lim_{q \searrow 0} \frac{C_{LU}(q, q)}{q}
\end{aligned}$$

$$\begin{aligned}
\lambda_{UL} &= 1 - \lim_{q \searrow 1} \frac{C(q, 1 - q)}{1 - q} \\
&= \lim_{q \searrow 1} \frac{(1 - q) - C(q, 1 - q)}{1 - q} \\
&= \lim_{q \searrow 1} \frac{C_{UL}(1 - q, 1 - q)}{1 - q} \\
&= \lim_{q \searrow 0} \frac{C_{UL}(q, q)}{q}
\end{aligned}$$

Por lo que, dado un vector aleatorio  $(X_1, X_2)$  con cópula  $C(u, v)$ , la medida total de dependencia del vector puede ser expresada en forma de matriz  $\Lambda$ :

$$\Lambda = \begin{pmatrix} \lambda_{LU} & \lambda_U \\ \lambda_L & \lambda_{UL} \end{pmatrix}$$

$$\Lambda = \lim_{q \rightarrow 0} \begin{pmatrix} \frac{C_{LU}(q, q)}{q} & \frac{\hat{C}(q, q)}{q} \\ \frac{C(q, q)}{q} & \frac{C_{UL}(q, q)}{q} \end{pmatrix}$$

La generalización de la función de dependencia extrema es:

$$\begin{aligned}
\Lambda_U(x_1, x_2) &= \lim_{t \rightarrow \infty} \frac{\mathbb{P}(X_1 > F_1^{-1}(1 - \frac{x_1}{t}), X_2 > F_2^{-1}(1 - \frac{x_2}{t}))}{\mathbb{P}(X_2 > F_2^{-1}(1 - \frac{x_2}{t}))} \\
&= \lim_{t \rightarrow \infty} \frac{1 - \mathbb{P}(X_1 \leq F_1^{-1}(1 - \frac{x_1}{t})) - \mathbb{P}(X_2 \leq F_2^{-1}(1 - \frac{x_2}{t})) + \mathbb{P}(X_1 \leq F_1^{-1}(1 - \frac{x_1}{t}), X_2 \leq F_2^{-1}(1 - \frac{x_2}{t}))}{1 - \mathbb{P}(X_2 \leq F_2^{-1}(1 - \frac{x_2}{t}))} \\
&= \lim_{t \rightarrow \infty} \frac{1 - (1 - \frac{x_1}{t}) - (1 - \frac{x_2}{t}) + C(1 - \frac{x_1}{t}, 1 - \frac{x_2}{t})}{\frac{x_2}{t}} \\
&= \lim_{t \rightarrow \infty} \frac{\hat{C}(\frac{x_1}{t}, \frac{x_2}{t})}{\frac{x_2}{t}} \\
&= \lim_{u \rightarrow 0} \frac{\hat{C}(ux_1, ux_2)}{ux_2}
\end{aligned}$$

$$\begin{aligned}
\Lambda_L(x_1, x_2) &= \lim_{t \rightarrow \infty} \frac{\mathbb{P}(X_1 \leq F_1^{-1}(\frac{x_1}{t}), X_2 \leq F_2^{-1}(\frac{x_2}{t}))}{\mathbb{P}(X_2 \leq F_2^{-1}(\frac{x_2}{t}))} \\
&= \lim_{t \rightarrow \infty} \frac{\mathbb{P}(X_1 \leq F_1^{-1}(\frac{x_1}{t}), X_2 \leq F_2^{-1}(\frac{x_2}{t}))}{1 - \mathbb{P}(X_2 > F_2^{-1}(\frac{x_2}{t}))} \\
&= \lim_{t \rightarrow \infty} \frac{C(\frac{x_1}{t}, \frac{x_2}{t})}{\frac{x_2}{t}} \\
&= \lim_{u \rightarrow 0} \frac{\hat{C}(ux_1, ux_2)}{ux_2}
\end{aligned}$$

#### 5.4. Resumen del capítulo

- La dependencia de cola es una medida utilizada para describir la magnitud de los valores que aparecen con variaciones en la otra variable. En la tesis abordamos coeficientes y funciones de dependencia extrema.
- Las funciones y coeficientes de dependencia extrema pueden escribirse en términos de cópulas.

### 6. Estimación y ajuste de cópulas y funciones de dependencia extrema

Considere una muestra aleatoria  $(X_{11}, X_{21}), \dots, (X_{1n}, X_{2n})$  de una función de distribución bivariada  $F$ . Si pensamos en el Teorema de Sklar, éste sugiere romper el modelo bivariado en dos partes independientes: el ajuste de las funciones de distribución marginales y la calibración de un modelo de cópula paramétrico apropiado.

## 6.1. Cópula empírica

Fueron estudiadas originalmente por Deheuvels en 1979. La idea consiste en construir una función cópula representando una frecuencia observada a partir de valores muestrales obtenidos de las variables univariadas sin establecer dependencia de ningún parámetro. De esta manera, la cópula es no paramétrica y queda definida únicamente a partir de la muestra.

**Definición 15.** Sea  $\{(x_k, y_k)\}_{k=1}^n$  una muestra de tamaño  $n$  de una distribución bivariada continua. La cópula empírica es la función  $C_n$  dada por:

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

Donde  $x_{(i)}$  y  $y_{(j)}$ ,  $1 \leq i, j \leq n$  denotan los estadísticos de orden de la muestra.

Otra expresión para la cópula empírica es la siguiente:

$$C_n \left( \frac{i}{n}, \frac{j}{n} \right) = \frac{1}{n} \sum_{k=1}^n \mathbb{I}(x_k \leq x_{(i)}, y_k \leq y_{(j)})$$

para  $j, k = 1, \dots, n$

Deheuvels también demostró que cuando  $n$  crece, entonces  $C_n \rightarrow C$ . De esta manera, la cópula empírica es un estimador consistente de la cópula verdadera.

## 6.2. Estimación de los parámetros de la cópula por el método de Máxima verosimilitud

La estimación de los parámetros de la cópula pueden hacerse por el método de la máxima verosimilitud completa o semi paramétrica dependiendo de la distribución disponible de las distribuciones marginales. Bajo el enfoque semi paramétrico, primero se sustituyen las distribuciones marginales por sus contrapartes empíricas y después se estiman los parámetros de la cópula por Máxima verosimilitud.

Sea  $X = (X_1, X_2)$  un vector aleatorio bidimensional con función de distribución conjunta  $F$  y marginales univariadas  $F_1(x_1)$  y  $F_2(x_2)$ .

Además suponga que la cópula pertenece a una familia paramétrica  $C = \{C_\theta : \theta \in \Theta\}$ .

Del teorema de Sklar sabemos que la distribución de  $X$  puede ser expresada como:

$$F(x_1, x_2) = C \{F_1(x_1), F_2(x_2); \theta\}$$

y su densidad como:

$$f(x_1, x_2; \theta) = c \{F_1(x_1), F_2(x_2); \theta\} f_1(x_1) f_2(x_2)$$

donde  $c(u_1, u_2) = \frac{\partial^2 C(u_1, u_2)}{\partial u_1 \partial u_2}$ .

Para una muestra de observaciones  $\{x_t\}_{t=1}^T$ ,  $x_t = (x_{1,t}, x_{2,t})^t$  y  $\theta$  la función de verosimilitud está dada por:

$$L(\theta; x_1, \dots, x_T) = \prod_{t=1}^T f(x_{1,t}, x_{2,t}; \theta)$$

y la función de logverosimilitud como:

$$l(\theta; x_1, x_2, \dots, x_T) = \sum_{t=1}^T \text{lnc} \{F_1(x_1), F_2(x_2); \theta\} + \sum_{t=1}^T (\ln f_1(x_{1,t}) + \ln f_2(x_{2,t}))$$

Note que la log verosimilitud puede ser dividida en dos partes, la primera involucra a la estructura de dependencia y la segunda a las densidades marginales.

Máxima verosimilitud canónica: en este método, las distribuciones marginales univariadas son estimadas a través de la función de distribución empírica  $\hat{F}$ . Para  $j = 1, 2$ :

$$\hat{F}_j(x) = \frac{1}{T+1} \sum_{t=1}^T \mathbb{1}(X_{j,t} \leq x)$$

Utilizando este método la pseudo verosimilitud queda como:

$$l(\theta) = \sum_{t=1}^T \text{lnc} \{F_{X_1}(x_1, \delta_1), F_{X_2}(x_2, \delta_2); \theta\}$$

y el parámetro estimado de la cópula está dado por:

$$\theta_{MVC} = \text{argmáx}_{\theta} l(\theta)$$

Es importante notar que en este método no necesita de alguna especificación de las marginales. Por lo anterior importante ya que es una alternativa que no implica una distribución a priori de las marginales.

### 6.3. Bondad de ajuste: Crámer von Mises (CvM)

Un aspecto en el que se debe poner atención especial es en si la cópula desconocida pertenece o no a una familia paramétrica elegida. De manera más formal, uno quiere probar que:

$$H_0 = C \in \mathbf{C} = \{C_{\theta} : \theta \in \Theta\} \text{ vs } H_1 : C \notin \mathbf{C}$$

Donde  $\theta$  es un subconjunto abierto de  $\mathbb{R}^q$  para algún entero  $q \geq 1$ .

El criterio de bondad de ajuste es el más poderoso para decidir si rechazar o aceptar cópulas paramétricas. Este tipo de pruebas utilizan la distancia entre el estimado y la llamada cópula empírica. La cópula empírica es un estimador consistente de la cópula desconocida  $C$ , sin importar que  $H_0$  sea cierta o no.

La distancia entre el estimado y la cópula empírica se obtienen usando la distancia de Crámer von Mises y la cópula más cercana a la cópula empírica es considerada como la que provee el ajuste más apropiado a la estructura de dependencia.

Para el caso bivariado, el procedimiento de la prueba puede ser resumido como sigue:

1. Obtenga su muestra aleatoria  $(z_1, \dots, z_n)$ , convirtiendo su muestra  $(x_1, \dots, x_n)$  en rangos normalizados<sup>3</sup>
2. Estime los parámetros por MV
3. Calcule la cópula empírica
4. Si tiene una expresión analítica para  $C_\theta$ , calcule la estadística CvM, con  $C(z_1, \dots, z_n)$  y  $C_{\hat{\theta}}(z_1, \dots, z_n)$

Como sugirieron Genest y Rémillard, una prueba de bondad de ajuste natural consiste en comparar  $C_n$  con una estimación  $C_{\theta_n}$  asumiendo que  $C \in \mathcal{C}$ .

De acuerdo a las grandes simulaciones de Genest et al (2009), la versión más poderosa de este proceso está basada en la estadística de Crámer von Misses:

$$S_n = \sum_{i=1}^n \{C_n(u_{i,n}, v_{i,n}) - C_{\theta_n}(u_{i,n}, v_{i,n})\}^2$$

#### 6.4. Estimación no paramétrica de la función de dependencia extrema

Suponga que  $(X_1, X_2), (X_1^{(1)}, X_2^{(1)}), \dots, (X_1^{(m)}, X_2^{(m)})$  son vectores aleatorios independientes e idénticamente distribuidos con función de distribución  $F$ , y marginales  $G, H$  y cópula  $C$ .

Note primero que la estimación no paramétrica es la más apropiada para la función de dependencia de cola ya que no hay parametrización alguna previa de la cópula de cola. La cópula empírica se define como:  $C_m(u, v) = F_m(G_m^{-1}(u), H_m^{-1}(v))$ ,  $(u, v) \in [0, 1]^2$  con  $F_m, G_m, H_m$  las funciones de distribución empíricas de  $F, G, H$ . La elección de las distribuciones empíricas para modelar las distribuciones marginales elimina cualquier confusión debida a ajustes paramétricos incorrectos en el ajuste de las marginales.

Sea  $R_{m1}^{(j)}$  y  $R_{m2}^{(j)}$  los rangos de  $X_1^{(j)}, X_2^{(j)}$ ;  $j = 1, \dots, m$  respectivamente. Una propuesta de estimadores es:

$$\begin{aligned} \hat{\Lambda}_{L,m}(x_1, x_2) &= \frac{m}{k} C_m\left(\frac{kx_1}{m}, \frac{kx_2}{m}\right) \\ &\approx \frac{1}{k} \sum_{j=1}^m \mathbb{I}_{R_{m1}^{(j)} \leq kx_1, R_{m2}^{(j)} \leq kx_2} \end{aligned}$$

$$\hat{\Lambda}_{U,m}(x_1, x_2) \approx \frac{1}{k} \sum_{j=1}^m \mathbb{I}_{R_{m1}^{(j)} > m-kx_1, R_{m2}^{(j)} > m-kx_2}$$

---

<sup>3</sup>Es decir, obtenga la distribución empírica de los datos

Para alguna  $k \in \{1, \dots, m\}$  elegida por el estadístico. Por resultados asintóticos supondremos que  $k = k(m) \rightarrow \infty$ , es decir  $k$  está en función de  $m$  y que  $\frac{k}{m} \rightarrow 0$  cuando  $m \rightarrow \infty$ .

A estos estimadores se les conoce como cópulas de cola empíricas.

Los lados derechos de los estimadores tienen 2 aproximaciones de los rangos de los estadísticos de orden que están basados en una pequeña modificación de la cópula empírica propuesta por Génest et al:

$$\hat{C}_m(u, v) = \frac{1}{m} \sum_{j=1}^m \mathbb{I}_{\{G_m(X_1^{(j)}) \leq u, H_m(X_2^{(j)}) \leq v\}}; (u, v) \in [0, 1]^2$$

Para estimar los coeficientes de dependencia extrema, note que en general, la estimación está dada por:

$$\lambda = \frac{m}{k} C_m\left(\frac{k}{m}, \frac{k}{m}\right)$$

## 6.5. Resumen del capítulo

- En 1979, Deheuvels enunció una forma de estimar a la cópula por la hoy llamada cópula empírica. Recordando que la cópula es una función de distribución es lógico pensar en la función de distribución empírica como estimador para la cópula. Además, Deheuvels demostró que cuando  $n$  crece, la cópula empírica tiende a la cópula verdadera.
- Dependiendo de la información disponible así como de la finalidad del estudio, se pueden estimar los parámetros de la cópula por máxima verosimilitud completa o semi-paramétrica. Una de las ventajas que tiene la estimación semi-paramétrica es que no se necesita ningún supuesto para la distribución de las marginales, pues estas son sustituidas por sus correspondientes funciones de distribución empírica. En la aplicación mostrada más adelante utilizaremos la estimación semi-paramétrica
- Una vez que hemos ajustado los parámetros de la cópula es necesario conocer si la cópula seleccionada es la que mejor ajusta a nuestros datos, entre las pruebas que existen para probar la bondad de ajuste de la cópula está la estadística Crámer von Misses que calcula la suma de los cuadrados de las distancias entre la cópula empírica y la cópula elegida.
- La función de dependencia extrema es la generalización de los coeficientes de dependencia extrema expresada como un límite. Por su definición es difícil encontrar un estimador adecuado y mucho se ha escrito al respecto. Aquí se aborda como un estimador a la cópula empírica evaluada en puntos que cumplen con características asintóticas.

## 7. Simulación de cópulas

A continuación se presentan los pasos necesarios para implementar simulaciones Monte Carlo de pseudo y no pseudo variables aleatorias que tienen cierta cópula  $C$ .

Lo que se hará es dar un algoritmo que nos permitirá simular un solo vector  $\mathbf{U} = (U_1, \dots, U_d)$  que tiene marginales uniformes y la cópula deseada.

Luego se pueden obtener marginales arbitrarias utilizando la transformación del cuantil. Si se repite el algoritmo  $n$  veces se obtiene una muestra de  $n$  pseudo variables aleatorias independientes e idénticamente distribuidas que admiten la cópula deseada.

### 7.1. Cópula Gaussiana y t-Cópula

El primer paso consiste en simular las pseudo variables. de acuerdo con la distribución multivariada subyacente y el segundo es transformar estos resultados a marginales uniformes utilizando la transformación del cuantil. Lo anterior se explica en el siguiente algoritmo:

*Algoritmo para la cópula Gaussiana Multivariada*

1. Para una matriz de covarianza arbitraria  $\Sigma$ , obtener la matriz de correlación tal que la matriz sea de  $n \times n$  con 1 en la diagonal y  $\rho$  en cualquier otro caso.
2. Obtener la descomposición de Cholesky  $\Sigma = A'A$
3. Generar pseudo v.a's.  $\tilde{X}_d$
4. Calcular  $(X_1, \dots, X_d)' = \mathbf{X} = A \tilde{X}$  de  $\tilde{X}(\tilde{X}_1, \dots, \tilde{X}_d)$
5. Realizar la transformación  $U_i = \phi(X_i), i = 1, \dots, d$  donde  $\phi$  es la función de distribución acumulativa de una normal estándar

*Algoritmo para la cópula t Multivariada*

1. Para una matriz arbitraria de covarianzas  $\tilde{\Sigma}$  obtener la matriz de correlación como en el paso 1 del algoritmo para la simulación de la cópula normal.
2. Generar normales multivariadas  $\mathbf{X}$  con matriz de covarianza determinada en el paso anterior.
3. Generar variables independientes  $\Psi \sim \chi_n^2$ , utilizando por ejemplo  $\psi = \sum_{i=1}^n Y_i^2$ , donde  $Y_i$  es  $N(0, 1)$
4. Realizar la transformación  $U_i = t_\nu(X_i/\sqrt{\psi/\nu}), i = 1, \dots, d$  donde  $t_\nu$  es la función de distribución de una t-distribución univariada con  $\nu$  grados de libertad

## 7.2. Cópulas arquimedianas

Las cópulas Gumbel, Clayton y Frank son un ejemplo de las cópulas con transformada de Laplace arquimediana (o TL Cópulas arquimedianas). Para esta clase de cópula, la inversa de la generadora  $\phi$  tiene una representación como transformada de Laplace de alguna función  $G$ .

Considere una función de distribución acumulada  $G$  y denote a su transformada de Laplace como:  $\tilde{G}(t) := \int_{x=0}^{\infty} \exp(-tx)dG(x)$ ,  $t \geq 0$

Sea  $\tilde{G}(t)(\infty) := 0$  y notemos que  $\tilde{G}$  es una función continua y estrictamente decreciente, lo que nos hace considerarla un buen candidato para  $\phi^{-1}$ . De hecho, generar una pseudo variable  $V$  con función de distribución acumulativa  $G$  y pseudo variables uniformes estándar iid  $X_1, \dots, X_d$  (que también dependen de  $V$ ). Sea::

$$U_i := \tilde{G}\left(-\frac{\ln X_i}{V}\right)$$

entonces el vector  $U$  tiene la estructura de cópula arquimediana deseada con generador  $\phi = G^{-1}$ .

Lo anterior se consolida en el siguiente algoritmo:

*TL Cópula arquimediana*

1. Generar una pseudo v.a.  $V$  con función de distribución acumulativa  $G$ 
  - Para una cópula Clayton,  $V$  se distribuye Gamma,  $Ga(1/\theta, 1)$  y  $\tilde{G}(t) = (1+t)^{-1/\theta}$
  - Para una cópula Gumbel,  $V$  se distribuye estable,  $St(1/\theta, 1, \gamma, 0)$  con  $\gamma = (\cos(\pi/2\theta))^\theta$  y  $\tilde{G}(t) = \exp(-t^{1/\theta})$
  - Para una cópula Frank,  $V$  es discreta con  $P(V = k) = (1 - e^{-\theta})^k / (k\theta)$  para  $K = 1, 2, \dots$
2. Generar pseudo v.a. i.i.d's. uniformes  $X_1, \dots, X_d$
3. Aplicar la transformación  $U_i = \tilde{G}\left(-\frac{\ln X_i}{V}\right)$ ,  $i = 1, \dots, d$

## 7.3. Cópula de Marshall-Olkin

Se debe recordar que las variables aleatorias  $X_1$  y  $X_2$  con cópula Marshall-Olkin se obtienen de v.a's. distribuidas exponencial  $Z_1, Z_2$  y  $Z_{12}$  con intensidades  $\lambda_1, \lambda_2$  y  $\lambda_{12}$ , respectivamente, por:

$$X_i = \min\{Z_i, Z_{12}\} \text{ con } i=1,2$$

Las marginales de  $X_i$  se calculan de manera sencilla y se obtiene que:

$$F_i(x) := \mathbb{P}(X_i \geq x) = 1 - \exp(-(\lambda_i + \lambda_{12})x).$$



Finalmente, obtenemos las v.a. distribuidas uniformemente estableciendo  $U_i = F_i(x_i)$  y  $U_1, U_2$  tiene la cópula deseada. Lo anterior se sintetiza en el siguiente algoritmo:

*Algoritmo para la cópula Marshall-Olkin*

1. Generar pseudo v.a.'s.  $Z_i \sim \exp(\lambda_i)$ ,  $i=1, \dots, 3$
2. Realizar la transformación  $U_i = 1 - \exp\{-(\lambda_i + \lambda_{12})\max\{Z_i, Z_3\}\}$ ,  $i = 1, 2$

Si se utiliza la representación vía  $\alpha_1, \alpha_2$  como en:

$$C_{\alpha_1, \alpha_2}(u_1, u_2) = \min\{u_2 u_1^{1-\alpha_1}, u_1 u_2^{1-\alpha_2}\}$$

se eligen para algún  $\lambda_{12}$  los parámetros  $\lambda_1$  y  $\lambda_2$ , tales que:

$$\alpha_i = \frac{\lambda_i \lambda_3}{\lambda_i + \lambda_3}, i = 1, 2$$

## 7.4. Cópulas en R

### Simulación de las principales cópulas en R

library(copula)

- Simulación de la cópula Gaussiana  

```
set.seed(1)
nor.cop=normalCopula(0.5)
x = rcopula(nor.cop,1000)
plot(x,main="Ejemplo de la cópula Normal")
persp(nor.cop,dcopula,main="Función de distribución de la cópula normal", col="lightblue")
persp(nor.cop,pcopula,main="Función de distribución de la cópula normal", col="green")
```
- Simulación de la cópula t-student  

```
t.cop=tCopula(c(0.5,0.3),dim=3,dispstr = "toep",df=2)
scatterplot3d(rcopula(t.cop,1000))
```
- Cópula de Frank en R  

```
frank.cop = frankCopula (2 , dim = 3)
scatterplot3d ( rcopula (frank.cop , 1000) )
frank.cop = frankCopula (2)
persp (frank.cop , dcopula ,main = "Densidad de la cópula de Frank ", col = "green3")
persp (frank.cop , pcopula , col = "lightblue ", main="Función de distribución de la cópula de Frank ")
```

```
clayton.cop = archmCopula ("frank", 2)
contour (frank.cop , dcopula , main= "Curvas de nivel de densidad de la c3pula de Frank")
```

- C3pula de Gumbel en R

```
gumbel.cop =gumbelCopula (4 , dim = 3)
scatterplot3d (rcopula(gumbel.cop , 1000))
gumbel.cop = gumbelCopula (1,0)
persp(gumbel.cop , dcopula ,main = "Densidad de la c3pula de Gumbel ", col = "green3
")
persp(gumbel.cop , pcpula , col ="lightblue ", main= "Funci3n de distribuci3n de la
c3pula Gumbel")
gumbel.cop =archmCopula ("gumbel", 5)
contour(gumbel.cop,dcopula)
```

- Copula de Clayton

```
clayton.cop=claytonCopula ( 2 , dim = 3)
scatterplot3d(rcopula(clayton.cop , 1000))
clayton.cop=claytonCopula (0.1)
persp(clayton.cop,dcopula,main = "Densidad de la c3pula de Clayton ",col = "green3 ")
clayton.cop = archmCopula ("clayton",2)
contour(clayton.cop , dcopula , main = "Curvas de nivel de la c3pula de Clayton
```

## 7.5. Resumen del cap3tulo

- Para la simulaci3n de c3pulas se procede de manera contraria que con la estimaci3n, ya que primero se simula la dependencia y luego las marginales. Lo anterior se debe a que al utilizar las c3pulas se puede modelar de manera independiente la estructura de dependencia de las marginales.

## 8. Fundamentos del An3lisis de series de tiempo financieras

En esta secci3n abordaremos temas relevantes para el manejo de datos financieros. Comenzaremos con un resumen de los modelos de series de tiempo m3s relevantes y continuaremos con una breve introducci3n a los hechos estilizados.

### 8.1. Series de tiempo financieras

Revisaremos conceptos esenciales en el an3lisis de series de tiempo tales como estacionariedad, auto-correlaci3n y su estimaci3n, ruidos blancos y procesos ARMA (Modelos auto-

regresivos de media móvil). Seguiremos con los modelos ARCH y GARCH (modelos generalizados auto-regresivos condicionalmente heterocedásticos) utilizados para capturar el importante fenómeno de la volatilidad.

En esta sección daremos un breve resumen de las bases para el análisis de las series de tiempo univariadas, poniendo especial atención en lo que es relevante para la modelación de los factores de riesgo de las series de rendimientos.

**Definición 16.** (*Serie de tiempo*) Una serie de tiempo es un proceso estocástico  $(X_t)_{t \in \mathbb{Z}}$ , es decir, una familia de variables aleatorias, indexadas por los enteros y definidas en algún espacio de probabilidad  $(\Omega, \mathcal{F}, \mathbb{P})$ .

*Momentos de una serie.* Asumiendo que existen, definimos:

- La función media  $\mu(t)$ :

$$\mu(t) = \mathbb{E}(X_t) \quad t \in \mathbb{Z}$$

- La función varianza  $var(t)$ :

$$var(t) = \mathbb{E}((X_t - \mu(t))^2) \quad t \in \mathbb{Z}$$

- La función de auto-covarianza  $\gamma(t, s)$  de  $(X_t)_{t \in \mathbb{Z}}$  es una función que describe las covarianzas entre las variables del proceso en cada par de instantes:

$$\gamma(t, s) = Cov(X_t, X_s) = \mathbb{E}((X_t - \mu(t))(X_s - \mu(s))) \quad t, s \in \mathbb{Z}$$

Por su definición, es claro que la función de auto-covarianza satisface que:  $\gamma(t, s) = \gamma(s, t)$  para toda  $t, s$  y que  $\gamma(t, t) = var(X_t)$ .

- La función de auto-correlación  $\rho(t, s)$  de  $(X_t)_{t \in \mathbb{Z}}$  es una función que describe las auto-correlaciones entre las variables del proceso en un par de instantes  $t, s \in \mathbb{Z}$  cualesquiera:

$$\rho(t, s) = \frac{\gamma(t, s)}{\sigma_t \sigma_s} \quad t, s \in \mathbb{Z}$$

*Estacionariedad* De manera general se habla de dos tipos de estacionariedad: fuerte y débil. Informalmente, decimos que una serie de tiempo es estrictamente estacionaria si sus propiedades estadísticas permanecen sin cambio en el tiempo, es decir, son las mismas que las del proceso  $(X_{t_0+t})_t$  para cualquier elección de  $t_0$ .

Formalmente, tenemos la siguiente definición:

**Definición 17.** (*Estacionariedad estricta, también llamada fuerte*) La serie de tiempo  $(X_t)_{t \in \mathbb{Z}}$  es estrictamente estacionaria si

$$(X_{t_1}, \dots, X_{t_n}) \stackrel{d}{=} (X_{t_1+k}, \dots, X_{t_n+k})$$

Con  $t_1, \dots, t_n, k$  en  $\mathbb{Z}$  y para toda  $n \in \mathbb{N}$ .

De manera informal, la noción de estacionariedad débil hace referencia a una serie de tiempo cuya función de media es constante y su función de auto-covarianza es también estable, formalmente:

**Definición 18.** (*Estacionariedad débil*) La serie de tiempo  $(X_t)_{t \in \mathbb{Z}}$  es débilmente estacionaria si:

- $\mathbb{E}(X_t^2) < \infty \forall t \in \mathbb{Z}$
- $\mu(t) = \mu \forall t \in \mathbb{Z}$
- $\gamma(t, s) = \gamma(t + k, s + k) \forall t, s, k \in \mathbb{Z}$

Ambas definiciones intentan formalizar la noción de que el comportamiento de una serie de tiempo es similar en cualquier ventana de tiempo en la que sea observado. Es importante notar que toda serie estrictamente estacionaria con varianza finita es débilmente estacionaria. El regreso no es cierto.

*Auto-correlación en series de tiempo estacionarias* La definición de estacionariedad débil implica que para toda  $s, t$  tenemos que  $\gamma(t - s, 0) = \gamma(t, s) = \gamma(s, t) = \gamma(s - t, 0)$ , con lo que la covarianza entre  $X_t$  y  $X_s$  sólo depende de su separación temporal  $|s - t|$ , que se conoce como el decaimiento. Así, para un proceso débilmente estacionario escribimos a la función de auto-covarianza como una función de una variable:

$$\gamma(h) := \gamma(h, 0), \forall h \in \mathbb{Z}$$

Al notar que  $\gamma(0) = \text{var}(X_t), \forall t$ , podemos definir la función de auto-correlación de un proceso débilmente estacionario:

**Definición 19.** (*Función de auto-correlación de un proceso débilmente estacionario*) La función de auto-correlación (ACF) denotada por  $\rho(h)$  de un proceso débilmente estacionario  $(X_t)_{t \in \mathbb{Z}}$  es:

$$\rho(h) = \rho(X_h, X_0) = \gamma(h)/\gamma(0), \forall h \in \mathbb{Z}$$

A continuación definiremos el concepto de ruido blanco, que es fundamental para crear series de tiempo útiles:

**Definición 20.** (*Ruido blanco*)  $(X_t)_{t \in \mathbb{Z}}$  es un ruido blanco si es débilmente estacionario con función de auto-correlación:

$$\rho(h) = \begin{cases} 1, & \text{si } h = 0, \\ 0, & \text{si } h \neq 0 \end{cases}$$

Un ruido blanco con media cero y varianza  $\sigma^2 = \text{var}(X_t)$  se denotará como  $RB(0, \sigma^2)$ . Un ejemplo simple de un proceso ruido blanco es una serie de variables aleatorias independientes e idénticamente distribuidas con varianza finita, y se conoce como un ruido blanco estricto.

### Modelos AR(p), MA(q), ARMA(p,q) y ARIMA(p,d,q)

**Definición 21.** (Proceso auto-regresivo de orden  $p$  AR(p)) Sea  $(W_t)_{t \in \mathbb{Z}}$  un  $RB(0, \sigma^2)$ . Una serie de tiempo con media cero  $(X_t)_{t \in \mathbb{Z}}$  se denomina auto-regresiva de orden  $p$  con respecto a  $(W_t)_{t \in \mathbb{Z}}$  si:

$$X_t = \phi_1 X_{t-1} + \dots + \phi_p X_{t-p} + W_t$$

para algunos  $\phi_1, \dots, \phi_p \in \mathbb{R}$

Una herramienta muy útil en la identificación del orden del modelo es la función de auto correlación parcial.

**Definición 22.** (Proceso de promedios móviles de orden  $q$  MA(q)) Sea  $(W_t)_{t \in \mathbb{Z}}$  un  $RB(0, \sigma^2)$ . Una serie de tiempo  $(X_t)_{t \in \mathbb{Z}}$  se denomina de promedios móviles de orden  $q$  con respecto a  $(W_t)_{t \in \mathbb{Z}}$  si:

$$X_t = W_t + \theta_1 W_{t-1} + \dots + \theta_q W_{t-q}$$

para algunos  $\theta_1, \dots, \theta_q \in \mathbb{R}$

Para conocer el orden del modelo MA, uno puede ayudarse de la función de auto correlación y auto covarianza, las cuales se anulan para los rezagos mayores al orden idóneo.

Debido a que el objetivo de este capítulo es introducir de manera básica ciertos modelos de series de tiempo, no se abordarán los temas de estimación y simulación para ninguno de los modelos. Sin embargo, el lector interesado puede encontrar en la bibliografía algunos libros especializados en el tema, por ejemplo [1].

**Definición 23.** (Proceso auto-regresivo de orden  $p$  y de medias móviles de orden  $q$  ARMA(p,q)). Sea  $(W_t)_{t \in \mathbb{Z}}$  un  $RB(0, \sigma^2)$ , el proceso  $(X_t)_{t \in \mathbb{Z}}$  es un proceso ARMA(p,q) de media cero con respecto a  $(W_t)_{t \in \mathbb{Z}}$  si es débilmente estacionario y satisface una ecuación de la forma:

$$X_t - \phi_1 X_{t-1} - \dots - \phi_p X_{t-p} = W_t + \theta_1 W_{t-1} + \dots + \theta_q W_{t-q}$$

para algunos  $\phi_1, \dots, \phi_p, \theta_1, \dots, \theta_q \in \mathbb{R}$

**Observación 2.**  $(X_t)_{t \in \mathbb{Z}}$  es un proceso ARMA con media  $\mu$  si la serie centrada  $(X_t - \mu)_{t \in \mathbb{Z}}$  es un proceso ARMA(p,q) de media cero.

De acuerdo con nuestra definición, se podría pensar que los procesos ARMA que no son débilmente estacionarios no existen. Sin embargo, el que el proceso sea estrictamente estacionario o no, dependerá de la naturaleza exacta del ruido blanco, conocido también como el proceso de innovaciones. Si las innovaciones son iid o forman por si mismas un proceso estrictamente estacionario, entonces el proceso ARMA lo será también.

*Causalidad.* Por motivos prácticos, restringiremos nuestro estudio al estudio de procesos ARMA a procesos ARMA causales. Por causales, entenderemos procesos que cumplen la igualdad de la definición anterior y que tienen una representación de la forma:

$$X_t = \sum_{i=0}^{\infty} \psi_i W_{t-i}$$

donde los coeficientes  $\psi_i$  deben satisfacer:

$$\sum_{i=0}^{\infty} |\psi_i| < \infty$$

De manera informal, la característica de causalidad nos habla de que los procesos pueden verse como función del pasado de un ruido blanco.

La función de auto-correlación de un modelo ARMA(p,q) está dada por:

$$\rho(h) = \frac{\sum_{i=0}^{\infty} \psi_i \psi_{i+|h|}}{\sum_{i=0}^{\infty} \psi_i^2}, h \in \mathbb{Z}$$

Dado que la propiedad de causalidad depende sólo de la parte auto-regresiva, entonces el proceso ARMA es un proceso causal si y sólo si el polinomio:

$$\phi(z) = 1 - \phi_1 z - \dots - \phi_p z^p$$

no tiene raíces en el círculo unitario.

*Invertibilidad.* De manera informal, si en un proceso ARMA(p,q) el RB( $W_t$ ) $_{t \in \mathbb{Z}}$  puede escribirse en términos de la serie  $X_t$  y sus valores pasados  $X_{t-1}, X_{t-2}, \dots$  llamamos invertibilidad a la condición de que el polinomio  $\theta(z)$  no tenga raíces en el círculo unitario.

A continuación definiremos al operador de retraso para después definir a los modelos Auto-regresivos Integrados de Medias Móviles:

**Definición 24.** *Operador de retraso*

Definimos a  $\nabla$  como el operador de retraso, de tal manera que para una serie de tiempo  $(Y_t)_{t \in \mathbb{Z}}$  tenemos  $\nabla Y_t = Y_t - Y_{t-1}$ .

Denotemos a las diferencias repetidas por  $\nabla^d$ , donde:

$$\nabla^d Y_t = \begin{cases} \nabla^d Y_t & , \text{ si } d = 1, \\ \nabla^{d-1}(\nabla Y_t), & \text{ si } d > 1. \end{cases}$$

**Definición 25.** (*Modelos Auto-regresivos Integrados de Medias Móviles*) La serie de tiempo  $(Y_t)_{t \in \mathbb{Z}}$  es un proceso ARIMA  $(p, d, q)$  si las series diferenciadas  $(X_t)_{t \in \mathbb{Z}}$  dadas por  $X_t = \nabla^d Y_t$  forman un proceso ARMA  $(p, q)$ .

Para  $d > 1$ , los procesos ARIMA son no estacionarios. En la práctica son frecuentemente utilizados ya que las diferencias pueden transformar datos no estacionarios a datos que pueden ser modelados como ARMA estacionarios.

**Análisis en el dominio de tiempo.** Supondremos que tenemos una serie de tiempo  $X_1, \dots, X_n$  débilmente estacionaria. El análisis en el dominio del tiempo involucra calcular estimados empíricos de auto-covarianzas y auto-correlaciones de esta muestra aleatoria y utilizarlos para hacer inferencia sobre la estructura de dependencia del proceso.

*Correlograma* Las auto-covarianzas muestrales se calculan de acuerdo a:

$$\gamma(h) = \frac{1}{n} \sum_{t=1}^{n-h} (X_{t+h} - \bar{X})(X_t - \bar{X}), 0 \leq h < n,$$

donde  $\bar{X} = \sum_{t=1}^n X_t/n$  es la media muestral, que estima a  $\mu$ , la media de la serie de tiempo.

Y como ya habíamos visto anteriormente, podemos ahora calcular la función de auto correlación muestral:

$$\hat{\rho}(h) = \hat{\gamma}(h)/\hat{\gamma}(0), 0 \leq h < n$$

El correlograma es la gráfica  $\{(h, \hat{\rho}(h)) : h = 0, 1, 2, \dots\}$  diseñada para facilitar la interpretación de la ACF muestral.

*Pruebas compuestas* Es frecuentemente útil combinar el análisis visual del correlograma con una prueba formal numérica de la hipótesis de ruido blanco estricto, y una prueba popular es la de Ljung-Box. Bajo la hipótesis nula de ruido blanco estricto, la estadística:

$$Q_{LB} = n(n+2) \sum_{j=1}^h \frac{\hat{\rho}(j)^2}{n-j}$$

tiene distribución asintótica ji-cuadrada con  $h$  grados de libertad. Si una serie de variables aleatorias forman un proceso ruido blanco estricto, entonces el valor absoluto o el cuadrado de las variables debe ser también independiente e idénticamente distribuido. Es una buena idea aplicar también el correlograma y la prueba Ljung-Box a los valores absolutos como una prueba más de la hipótesis de ruido blanco estricto.

### Análisis estadístico de una serie de tiempo

En la práctica, el análisis estadístico de una serie de tiempo  $X_1, \dots, X_n$  sigue los siguientes pasos:

- *Análisis preliminar.* Se grafican los datos y se considera el ajuste de un modelo estacionario. Ya que aquí nos concentraremos en el análisis de series de tiempo logarítmicas diferenciadas, supondremos que la información ha sido *manipulada* previamente. Si bien algunas series financieras presentan patrones estacionales supondremos que estos efectos son menores en series semanales o diarias. El uso de *high frequency data* puede tener como principal problema la limpieza de los datos debido a que estas series muestran ciclos diarios y otras características deterministas.

Es claro que la hipótesis de estacionariedad se vuelve más cuestionable si tomamos ventanas de tiempo más largas o si escogemos ventanas en las que políticas económicas bien conocidas tienen lugar. Es una buena idea realizar análisis con distintas ventanas del tiempo y medir la sensibilidad entre la inferencia estadística y la cantidad de datos.

- *Análisis en el dominio del tiempo* Se evalúa la hipótesis de estacionariedad estricta al aplicar correlogramas y pruebas compuestas tales como la Ljung-Box a los *datos duros* (datos originales, en el sentido de que no han sido transformados ni han sufrido alguna modificación) y a sus valores absolutos. Si no se rechaza la hipótesis, entonces se termina el análisis formal.

Para las series diarias esperamos rechazar la hipótesis de ruido blanco estricto. Apesar del hecho que los correlogramas de los *datos duros* pueden mostrar poca evidencia de correlación serial, los correlogramas de los valores absolutos más probablemente mostrarán evidencia de dependencia serial fuerte. Es decir, los datos soportan un modelo de ruido blanco pero no uno de ruido blanco estricto. En estos casos no se necesita un modelo ARMA pero un modelo para la volatilidad puede ser útil.

Si el correlograma provee evidencia de un tipo de patrones de correlación serial producidos por procesos ARMA, entonces podemos intentar el ajuste de un modelo ARMA a nuestros datos.

- *Ajuste del modelo.* De manera frecuente se suele identificar primero el orden de un probable modelo ARMA usando el correlograma y otras herramientas como por ejemplo el correlograma parcial. Con la ayuda de un software podemos ajustar diversos modelos y elegir el *mejor* con la ayuda de un criterio de selección de modelos como el Akaike (AIC).

Algunas veces existen razones apriori para esperar que ciertos tipos de modelos sean los más apropiados. Una vez que escogimos el modelo para ajustar, existen muchos métodos de ajuste, incluyendo modelos especializados para procesos AR como las ecuaciones de Yule Walker.

- *Análisis de residuales y comparación de modelos* En un modelo ARMA, los residuales deben comportarse como una realización de un proceso ruido blanco, debido a que ese es uno de los supuestos del modelo para las innovaciones, y puede ser evaluado construyendo el correlograma. Si aún hay evidencia de que existe correlación serial en



el correlograma entonces esto sugeriría que el modelo ARMA adecuado aún no ha sido hallado. Además, se pueden utilizar pruebas compuestas para comprobar de manera formal que los residuales se comportan como una realización de un proceso ruido blanco estricto. Si los residuales se comportan como RBE (ruido blanco estricto) entonces no se necesita más modelado de series de tiempo. Si se comporta como un ruido blanco pero no como un RBE, entonces un modelo de volatilidad será requerido.

Es posible encontrarse con más de un modelo ARMA razonable para los datos, y entonces se requerirá de técnicas formales de comparación de modelos para decidir cuál o cuáles son los mejores. El criterio Akaike puede ser usado, al igual que cualquiera de sus variantes.

- *Predicción* Existen diversos enfoques de pronóstico o predicción de series de tiempo. Mencionaremos dos que se extienden de manera sencilla al caso de los modelos GARCH. La primera estrategia usa los modelos ajustados ARMA (o ARIMA) y algunas veces es llamado enfoque Box-Jenkins. La segunda estrategia es un enfoque libre de modelo para pronosticar conocido como *suavizamiento exponencial*, que se encuentra relacionado con la técnica de los promedios móviles ponderados exponencialmente para predecir la volatilidad.

Como ya hemos mencionado, el objetivo de esta tesis no es ahondar mucho en detalles de series de tiempo sino más bien dar al lector una visión general de conceptos de este tema, por lo que se recomienda revisar la referencia [4] página 137 para más información referente a predicción.

## Modelos para volatilidad cambiante y aleatoria

La volatilidad es una característica inherente a las series de tiempo financieras. Mide el tamaño de los errores hechos en el modelado de los rendimientos y otras variables financieras. En general, no es constante y en consecuencia los modelos de series de tiempo tradicionales que suponen varianza homocedástica, no son adecuados para modelar series de tiempo financieras.

En 1982 Engle introduce una nueva clase de procesos estocásticos llamados modelos ARCH, en los cuales la varianza condicionada a la información pasada no es constante y depende del cuadro de las innovaciones pasadas. Bollerslev en 1986 generaliza los modelos ARCH al proponer los modelos GARCH en los cuales la estructura de la varianza condicional depende, además del cuadro de los errores retrasados  $q$  períodos como en el modelo ARCH( $q$ ), de las varianzas condicionales retrasadas  $p$  períodos.

## Modelo ARCH

Por sus siglas en inglés Autoregressive Conditional Heteroskedasticity o modelo Autoregresivo con Heterocedasticidad Condicional. Este modelo hace referencia a la varianza o volatilidad del error  $\epsilon_t$ .

Utiliza la siguiente notación:

$$\sigma_t^2 = \text{var}(\epsilon_t)$$

La varianza del error no es constante por lo que tenemos la propiedad de heterocedasticidad. El modelo ARCH con  $p$  retrasos se denota como  $ARCH(p)$ .

Lo que el modelo nos dice de manera informal es que la volatilidad actual es el promedio de los errores pasados al cuadrado:

$$\sigma_t^2 = \alpha_0 + \alpha_1 \epsilon_{t-1}^2 + \dots + \alpha_p \epsilon_{t-p}^2$$

No entraremos a gran profundidad en el tema de estimación de los modelos ARCH, solamente mencionaremos que podemos utilizar máxima verosimilitud (pero que también existen otros estimadores).

Este tipo de modelos tiene propiedades similares a los modelos AR solo que las propiedades en este caso se relacionan con la volatilidad de la serie. A continuación mostraremos la definición formal de este proceso:

**Definición 26.** Sea  $(Z_t)_{t \in \mathbb{Z}}$  un  $RBE(0,1)$ . El proceso  $(X_t)_{t \in \mathbb{Z}}$  es un proceso  $ARCH(p)$  si es estrictamente estacionario y si satisface que para toda  $t \in \mathbb{Z}$  y algún proceso estrictamente positivo  $(\sigma_t)_{t \in \mathbb{Z}}$  se cumplen las ecuaciones:

$$\begin{aligned} X_t &= \sigma_t Z_t \\ \sigma_t^2 &= \alpha_0 + \sum_{i=1}^p \alpha_i X_{t-i}^2, \end{aligned}$$

donde  $\alpha_0 > 0$  y  $\alpha_i \geq 0, i = 1, \dots, p$

## Modelo GARCH

Por sus siglas en inglés Generalized Autoregressive Conditional Heteroskedasticity o modelo Generalizado Autoregresivo con Heterocedasticidad Condicional. Es la extensión del modelo ARCH más popular. El modelo ARCH tiene una volatilidad que depende de los errores anteriores al cuadrado. El modelo GARCH le agrega retrasos de la volatilidad misma.

Las propiedades del modelo ARCH y del modelo GARCH son similares, aunque el modelo GARCH es más flexible ya que es capaz de *atrapar* diferentes patrones de la volatilidad financiera.

Este modelo se denota como  $GARCH(p,q)$  y tiene la siguiente ecuación de volatilidad:

$$\sigma_t^2 = \alpha_0 + \alpha_1 \epsilon_{t-1}^2 + \dots + \alpha_p \epsilon_{t-p}^2 + \beta_1 \sigma_{t-1}^2 + \dots + \beta_q \sigma_{t-q}^2$$

A continuación enunciaremos la definición formal de este proceso:

**Definición 27.** Sea  $(Z_t)_{t \in \mathbb{Z}}$  un  $RBE(0,1)$ . El proceso  $(X_t)_{t \in \mathbb{Z}}$  es un proceso  $GARCH(p)$  si es estrictamente estacionario y si satisface que para toda  $t \in \mathbb{Z}$  y algún proceso estrictamente positivo  $(\sigma_t)_{t \in \mathbb{Z}}$  se cumplen las ecuaciones:

$$X_t = \sigma_t Z_t$$

$$\sigma^2 = \alpha_0 + \sum_{i=1}^p \alpha_i X_{t-i}^2 + \sum_{j=1}^q \beta_j \sigma_{t-j}^2,$$

donde  $\alpha_0 > 0$ ,  $\alpha_i \geq 0, i = 1, \dots, p$  y  $\beta_j \geq 0, j = 1, \dots, q$

## 8.2. Hechos estilizados en datos financieros

Los hechos estilizados son aquellos que se obtienen de tomar un común denominador entre las propiedades observadas en estudios realizados sobre diversos mercados e instrumentos. Es claro que al hacer esto uno logra algunas generalizaciones pero pierde precisión de las afirmaciones que se pueden hacer sobre los rendimientos de los activos.

De hecho, los hechos estilizados están formulados en términos de propiedades cualitativas de las series y pueden no ser lo suficientemente precisos para distinguir entre diferentes modelos paramétricos.

Si el lector quisiera obtener más información sobre este tema, puede recurrir a las siguientes referencias: [2] y [4].

### Propiedades estadísticas estilizadas de series de rendimientos

A continuación enunciaremos una serie de propiedades estadísticas que se ha demostrado están presentes en los datos financieros de diversos mercados con frecuencia diaria. No hay una definición general de qué es un hecho estilizado, o cuántos hay, mostraremos algunos que son relevantes en el estudio de dependencia de cola en datos financieros que se aborda en esta tesis.

1. **Presencia de auto-correlación** Esto se contrapone a la hipótesis de variables aleatorias e idénticamente distribuidas que utiliza todo modelo de series de tiempo.
2. **Colas pesadas** Las series de rendimientos financieras suelen presentar colas pesadas. Esta característica se debe a los movimientos no predecibles del mercado, es decir, un gran número de valores extremos implica un nivel más alto de curtosis de aquel que presenta la distribución normal.
3. **Gaussianidad** Conforme se incrementa la escala de tiempo sobre la que se calculan los rendimientos, la distribución de los mismos se parece cada vez más a una normal. Lo anterior, se sustenta teóricamente en el Teorema Central del Límite.
4. **Intermitencia** Los rendimientos presentan, en cualquier escala del tiempo, un alto grado de variabilidad.

5. **Presencia de volatilidad** Los rendimientos suelen presentar en cualquier escala de tiempo cierto grado de varianza. Además, los conocidos como clústers (agrupamientos) de volatilidad en las series de rendimientos también suelen estar formados por eventos extremos de los datos. Esta propiedad se relaciona con la característica de colas pesadas que como vimos, también poseen este tipo de series.
6. **Presencia de volatilidad** El volumen de la transacción está correlacionado con las medidas de volatilidad.

### 8.3. Resumen del capítulo

- Una serie de tiempo es una familia de variables aleatorias, indexadas por los enteros y definidas en algún espacio de probabilidad. Existen dos métodos en el análisis de las series de tiempo: del dominio del tiempo y del dominio de frecuencias. Nosotros utilizaremos el primero, que estudia la auto-correlación de la serie.
- Para realizar nuestro análisis utilizando modelos ARIMA y GARCH debemos verificar que nuestras series son estacionarias, lo que de manera informal significa que sus propiedades estadísticas permanecen invariantes en el tiempo. Si nuestra serie no es estacionaria, debemos transformarla hasta obtener una que sí lo sea.
- Los modelos ARMA son utilizados para modelar la esperanza condicional de un proceso condicionado al pasado, sin embargo este tipo de modelos supone que la varianza condicional dado el pasado es constante. El modelo ARMA debe ser actualizado cada vez que se tengan nuevas observaciones así como la estimación de los parámetros.
- Por lo anterior, los modelos ARMA (o ARIMA, dependiendo de cómo lo escribamos) son utilizados de manera conjunta con los modelos GARCH que son utilizados para modelar el término de ruido de un modelo ARMA.
- Los hechos estilizados de series financieras son un término utilizado para referirse al conjunto de evidencias estadísticas no triviales que se observan en los mercados financieros, independientemente de qué tipo sea este y de la escala del tiempo en el que se encuentren los datos. En el análisis de dependencia entre variables financieras algunos de los hechos más importantes son: la presencia de colas pesadas, clústers de volatilidad, entre otras.

## 9. Cálculo del VaR utilizando funciones de dependencia extrema y cópulas

En este ejemplo práctico se calculará el VaR de un portafolio de dos activos utilizando cópulas. El algoritmo resumido es:

1. Ajustar un modelo ARMA-GARCH a los log-rendimientos<sup>4</sup> de los componentes de un portafolio financiero.
2. Ajustar distribuciones Pareto Generalizadas a cada serie de residuales y aplicar Proposición 3(2) para obtener las marginales de la cópula.<sup>5</sup>
3. Ajustar una cópula al ajuste anterior.
4. Realizar simulación Monte Carl<sup>6</sup> de la cópula ajustada. Repetir este paso  $n$  veces.
5. Regresar las simulaciones a los precios originales y calcular el VaR de esas observaciones

Sin duda, el punto más importante del algoritmo anterior es el ajuste de las marginales, debido a que la calidad del ajuste final depende de este punto. Una vez garantizado el ajuste de las marginales el ajuste de la cópula cobra importancia, ya que además de decidir que cópula usar se deben determinar valores específicos para los parámetros. Como se mencionó anteriormente, la estimación de los parámetros se puede realizar utilizando diversos enfoques dependiendo de la información disponible de las distribuciones marginales, entre ellos los más comunes son:

- Máxima verosimilitud paramétrica completa: requiere hacer algunos supuestos sobre las distribuciones marginales. Los vector de parámetros (de las marginales y de la cópula) se estiman en un *sólo paso* mediante la maximización de la función de verosimilitud.
- Máxima verosimilitud paramétrica paso a paso. En este enfoque se estiman primero los parámetros de las distribuciones marginales y se usan para estimar el vector de parámetros de la cópula en un segundo paso.
- Máxima verosimilitud semi paramétrica: las distribuciones marginales se sustituyen por sus contrapartes empíricas y finalmente se maximiza la función de máxima verosimilitud.<sup>7</sup>
- Utilizando las estadísticas de bondad de ajuste, minimizando la distancia entre la cópula empírica y la cópula propuesta. Existen diversas estadísticas para este fin, la más popular es la estadística de Crámer von Misses.

---

<sup>4</sup> $r_t = \log\left(\frac{P_t}{P_{t-1}}\right) = \log(P_t) - \log(P_{t-1})$ . Véase la introducción en la siguiente sección para mayor información.

<sup>5</sup>La razón por la que trabajamos con los residuales de las series y no con las series originales es porque para el ajuste de la cópula, necesitamos que nuestras observaciones sean variables aleatorias. Esto lo conseguimos ajustando los modelos ARMA-GARCH.

<sup>6</sup>Utilizamos simulación Monte Carlo porque aunque existen métodos analíticos para calcular el VaR para la distribución normal (en este caso, la cópula gaussiana), en general se tienen que utilizar métodos numéricos para el cálculo del VaR de otras distribuciones. Se necesitan generar parejas de variables aleatorias  $(X_1, X_2) \sim F^{(C)}$  que representen escenarios de posibles cambios en el factor del riesgo. En la simulación Monte Carlo, se generan  $n$  de esos escenarios y se evalúa el valor del portafolio bajo cada escenario para después obtener el VaR.

<sup>7</sup>Véase Sección 6.2 para mayor información al respecto.

En nuestro caso realizaremos lo siguiente: A cada uno de los residuales del ajuste ARIMA los utilizaremos para estimar los parámetros de la cópula y después se seguirán los puntos 2 y 3 del algoritmo.

Además, proponemos una modificación a la manera de cómo estimar la cópula. Utilizaremos un estimador de la función de dependencia extrema y obtendremos un valor para el mismo basándonos en nuestros datos y con ello estimaremos los parámetros de las cópulas. Después, utilizando las características de nuestros datos, propondremos posibles cópulas que nos ayuden a modelar la dependencia de los mismos.

## 9.1. Introducción

Antes de comenzar, es bueno hacer algunas aclaraciones. La primera de ellas es que trabajaremos con los log-rendimientos diarios en lugar de los precios. Existen diversas razones para explicar esto como el hecho de que las series se vuelven estacionarias. Aunque la forma en la que calculamos los rendimientos es bajo el supuesto de precios continuos, debemos recordar que ésta es una muy buena aproximación para el caso discreto.

Lo que claramente se busca al invertir, es obtener una ganancia relativamente más grande a la inversión inicial. El resultado de una inversión, ya sea pérdida o ganancia, depende del precio y cantidad del activo invertido. Los rendimientos sirven como indicador de qué tan grande o pequeña es la pérdida o ganancia de la inversión en un activo (acciones, bonos, portafolio de acciones, etc). Los rendimientos son cambios en el precio expresados como una fracción del precio inicial.

Existen diferentes tipos de rendimientos: netos, brutos o los llamados log-rendimientos. Para los fines de la presente tesis, nos enfocaremos únicamente en estos últimos. Los *log-rendimientos* también llamados rendimientos compuestos continuamente, son denotados por  $r_t$  y se definen como:

$$r_t = \log\left(\frac{P_t}{P_{t-1}}\right) = \log(P_t) - \log(P_{t-1})$$

donde  $P_t$  es el precio de un activo al tiempo  $t$ .

Los *log-rendimientos* son aproximadamente iguales a los rendimientos que se definen como:  $R_t = \frac{P_t - P_{t-1}}{P_{t-1}}$ , recordando que  $\log(1 + x) \approx x$  si  $x$  es pequeña (menor a 10%).

Es común encontrar trabajos que supongan que los rendimientos se distribuyen de manera normal y que los log rendimientos son mutuamente independientes. Aunque la densidad de los rendimientos tiene forma de campana como la distribución normal, las colas de las distribuciones de los log-rendimientos son generalmente mucho más pesadas. Una distribución  $t$  con un parámetro pequeño para los grados de libertad (4 o 6) proporciona un mejor ajuste

que un modelo normal.

La hipótesis de independencia tampoco se cumple, debido a la existencia de correlación entre los rendimientos aunque generalmente son pequeñas. Más importante es que los rendimientos exhiben clustering de volatilidad. El clustering de volatilidad puede detectarse viendo las correlaciones entre los rendimientos al cuadrado.

De manera resumida el procedimiento a seguir es el siguiente:

- Ajustar un modelo ARMA a cada una de las series de los log-rendimientos.
- Revisar los residuales de cada ajuste. Si es necesario, ajustaremos modelo GARCH a cada serie de residuales.<sup>8</sup>
- Estandarizar<sup>9</sup> las series de residuales:

$$Z_t = \epsilon_t / \sigma_t.$$

- Verificar si cada una de las series de los residuales estandarizados es independiente e idénticamente distribuida.
- Ajustar distribución Pareto Generalizada (PG)<sup>10</sup> a cada serie de residuales estandarizados.
- Aplicar la distribución ajustada en el paso anterior a cada serie de residuales de manera respectiva, y ajustar una cópula a estos datos.
- Simular un horizonte de corto plazo de la cópula ajustada en el paso anterior.
- Aplicar la inversa de la función de distribución PG para obtener residuales estandarizados simulados.
- Obtener valores de  $\sigma_t^2$  tantos como simulaciones tengamos y convertir los residuales estandarizados simulados a residuales simulados.
- Regresar del modelo ARMA a los precios originales.

---

<sup>8</sup>Revisar la gráfica de los residuales al cuadrado, así como el ACF de los mismos.

<sup>9</sup>Esto con la finalidad de incorporar el modelo GARCH a nuestro ajuste

<sup>10</sup>En general, podemos ajustar la distribución que deseemos, por ejemplo, la distribución empírica. En nuestro caso, elegimos la distribución Pareto Generalizada ya que ésta se suele utilizar cuando no se tienen muchos datos en las colas para robustecer el ajuste.

## 9.2. Valor en riesgo (VaR)

El mundo financiero siempre ha sido riesgoso. La inclusión de nuevos instrumentos de inversión cada vez más complejos han resaltado su importancia pero también han llevado su cálculo a niveles cada vez más altos de complejidad. Existen diferentes tipos de riesgo: de mercado, crediticio, de liquidez y operacional. En esta aplicación nos enfocaremos al riesgo de mercado que es debido a cambios en los precios de instrumentos financieros. Existen diversos intentos de cuantificar los riesgos como el Valor en Riesgo (VaR) y el Déficit Esperado (ES) que son ampliamente usados porque pueden ser aplicados a todo tipo de riesgo y valores, incluyendo portafolios complejos.

El VaR utiliza dos parámetros, el horizonte de tiempo y el nivel de confianza, los cuales son denotados por  $T$  y  $1 - \alpha$ . Considerando lo anterior, el VaR es una cota tal que la pérdida hasta el horizonte fijado es menor que esa cota con probabilidad igual al nivel de confianza. Generalmente se escribe  $VaR(T, \alpha)$  para enfatizar la dependencia del VaR en ambos parámetros.

Si  $\mathcal{L}$  es la pérdida en el período  $T$ , entonces  $VaR(\alpha)$  es el  $\alpha$ -ésimo cuantil superior de  $\mathcal{L}$ . De manera equivalente, si  $\mathcal{R} = -\mathcal{L}$  es la ganancia, entonces  $VaR(\alpha)$  es menos el  $\alpha$ -ésimo cuantil de  $\mathcal{R}$ . Para funciones de pérdida continuas, el  $VaR(\alpha)$  :

$$P(\mathcal{L} > VaR(\alpha)) = P(\mathcal{L} \geq VaR(\alpha)) = \alpha$$

Y para cualquiera que sea la distribución de la pérdida, continua o no, se tiene que:

$$VaR(\alpha) = \inf\{x : \mathbb{P}(\mathcal{L} > x) \leq \alpha\}$$

## 9.3. Cálculo del VaR de un portafolio

Cuando se calcula el  $VaR$  de un portafolio de activos en lugar de un activo único, la estimación paramétrica basada en la hipótesis de una distribución multivariada  $t$  o normal de los rendimientos es muy conveniente, porque los rendimientos del portafolio tendrán un rendimiento univariado normal o  $t$ .

Estimar el  $VaR(\alpha)$  se torna complicado cuando el portafolio se encuentra integrado por diferentes tipos de activos como pueden ser bonos, acciones, opciones, y otros activos. Sin embargo, cuando un portafolio contiene solo acciones, el  $VaR(\alpha)$  es relativamente directo de estimar. En esta tesis, nos enfocaremos en este caso únicamente y mencionaremos algunos libros donde pueden encontrarse ejemplos más complejos.

La simulación Montecarlo supone que los movimientos en los factores de riesgo son generados por una distribución estimada:

$$\delta f^k \sim g(\theta), k = 1, \dots, K$$



donde  $g$  es la distribución conjunta y  $\theta$  sus parámetros.

Supongamos que observamos datos desde el día 1 hasta hoy, y que  $r_t$  es el rendimiento del portafolio en el día  $t$ , entonces tenemos una serie de rendimientos  $\{R_{t+1-\tau}\}_\tau = 1^m$ . El valor en riesgo se calcula como:

$$VaR(\alpha) = \text{percentil}\{\{R_{t+1-\tau}\}_\tau = 1^m, (100 - \alpha \%)\}$$

A diferencia de otros métodos paramétricos, la simulación Monte Carlo no hace alguna suposición sobre la distribución de los rendimientos.

Como ya se ha mencionado, los rendimientos diarios de acciones presentan un pequeño grado de auto-correlación y uno mayor de clúster o agrupamiento de la volatilidad. Cuando se calculan medidas de dispersión, la auto correlación puede ser ignorada si es lo suficientemente pequeña, pero el agrupamiento de la volatilidad es más difícil de ignorar. Lo que haremos será utilizar modelos ARMA+GARCH para que el  $VaR(\alpha)$  pueda ser ajustado a períodos de alta o baja volatilidad.

#### 9.4. Descripción de las series

Hemos seleccionado un portafolio hipotético<sup>11</sup> compuesto por dos acciones: Bimbo (BIMBOA.MX) y Mexchem (MEXCHEM.MX). Dichas acciones son dos de las más comercializadas en el país, además de formar parte de la muestra para el cálculo del Índice de Precios y Cotizaciones (IPC) que es el principal indicador del Mercado Mexicano de Valores. Usaremos los precios al cierre de estas dos acciones para el período comprendido entre el 1 de enero del 2012 hasta el 31 de diciembre del 2015 (restringiendo las observaciones a los días hábiles del mercado mexicano para ambas series. En total, contaremos con 1044 valores por cada serie).

Los estadísticos principales de estas dos acciones son los siguientes:

Estadística	Mexchem	Rend.	Bimbo	Rend.
Mínimo	12.7	-0.06	59.14	-0.05
1er cuantil	14.3	-0.01	70.03	-0.01
Mediana	15.38	0	78.01	0
Media	15.53	0	79.53	0
3er cuantil	16.81	0.01	89.69	0.01
Máximo	17.82	0.05	104.73	0.05
Coficiente de asimetría	-0.13	-0.09	0.19	-0.01
Curtosis	1.68	4.04	1.87	4.36
Varianza	1.93	0	137.57	0

<sup>11</sup>En nuestro portafolio, cada acción tiene como ponderador  $\frac{1}{2}$

Lo primero que haremos será graficar los datos originales para detectar tendencias que podamos remover y analizar la estructura de los mismos. Es importante aclarar que las gráficas no consideran ajuste alguno por pago de dividendos.

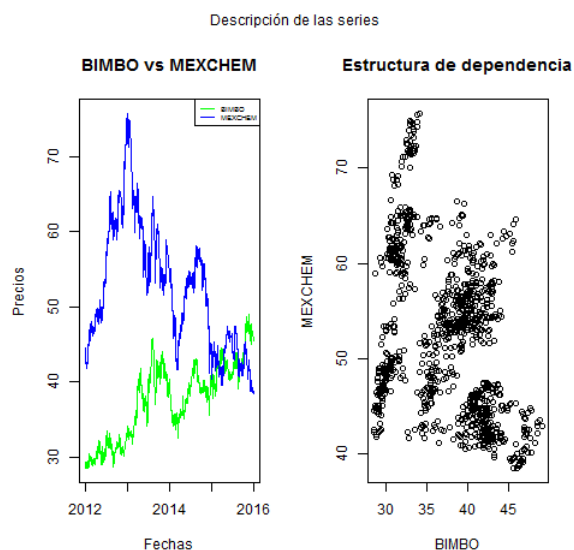


Figura 8: A la izquierda, la gráfica de los *datos duros* de ambas series. A la derecha, la gráfica de dispersión entre ambas series.

La gráfica anterior nos muestra que los datos tienen una tendencia muy similar en ciertos períodos del tiempo mientras que en otros, el comportamiento es totalmente diferente, por lo que la correlación entre ellos en diferentes ventanas del tiempo cambia de signo y de magnitud, lo que es claramente apreciable en la gráfica de la estructura de dependencia. Al calcularla para todo el período obtenemos un valor de  $-0.36$ , lo que nos indicaría que la relación lineal entre las dos series es negativa y medianamente baja.

Por lo anteriormente expuesto, realizaremos un análisis de la gráfica anterior fragmentando en períodos no necesariamente de la misma longitud pero en los que se aprecia un comportamiento similar en ambas series:

Además, revisando las gráficas de las series de manera individual podemos notar que no hay ningún patrón o tendencia aparente en el movimiento del precio de las acciones. Estas observaciones nos hacen pensar que una transformación logarítmica de los datos originales pueden ayudarnos a estabilizar la varianza.

A continuación veremos una comparación del cálculo de las medidas de asociación en cada una de las ventanas de tiempo:

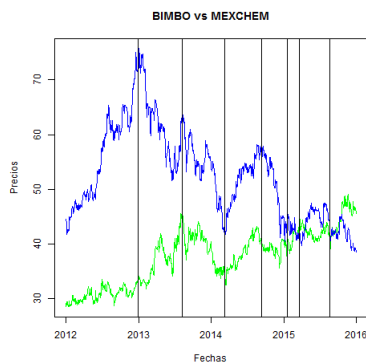


Figura 9: Gráfica de las series con el intervalo de fechas inicial fragmentado en los intervalos donde se observan patrones entre la relación de las series.

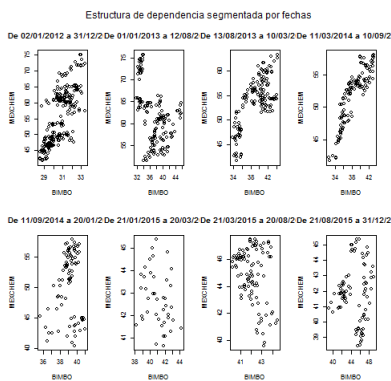


Figura 10: Gráfica de dispersión entre las series durante cada intervalo.

## 9.5. Ajuste del modelo ARIMA-GARCH

Comenzaremos con el ajuste de los modelos ARIMA-GARCH. Se recomienda primero ajustar el modelo ARIMA y después el modelo GARCH, de manera separada. Debido a que, como ya se mencionó, no se encontró ningún patrón lineal reconocible en los datos, calcularemos la primera diferencia de los logaritmos de los datos originales que, además de ayudarnos con la estabilización de la varianza tiene una interpretación en términos financieros (estamos calculando los llamados log-rendimientos).

Recordemos que para comenzar con el ajuste de un modelo ARMA nuestras series deben ser estacionarias. Para convertir series no estacionarias a estacionarias, calcular las diferencias de la serie puede ayudar.

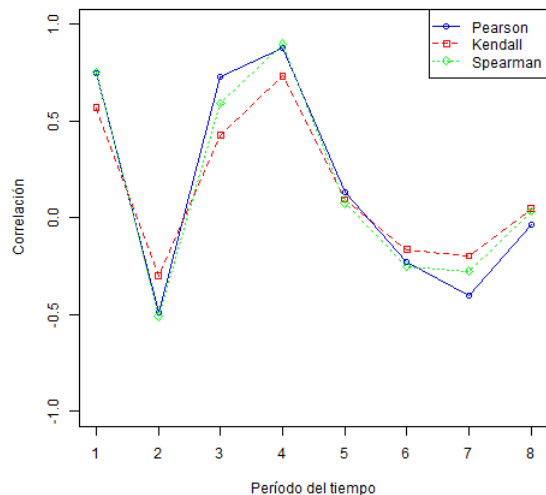


Figura 11: Medidas de dispersión y correlación de rangos en cada intervalo.

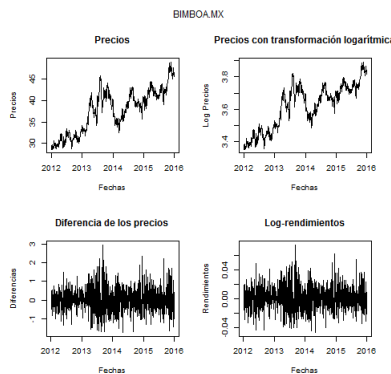


Figura 12: Gráficas de la serie BIMBO: Superior izquierda, la serie original de precios de la acción con un aparente crecimiento exponencial. Inferior izquierda, las diferencias de los precios de la acción. Podemos apreciar que la varianza de la serie incrementa cuando el nivel original de la serie incrementa, por lo que no es estacionaria. Superior derecha, la transformación logarítmica de la serie original. Inferior derecha, los rendimientos logarítmicos de la serie.

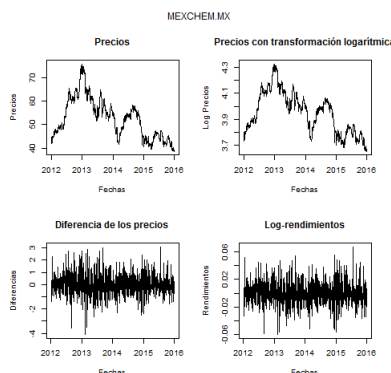


Figura 13: Gráficas de la serie MEXCHEM: Superior izquierda, la serie original de precios de la acción con un aparente crecimiento exponencial. Inferior izquierda, las diferencias de los precios de la acción. Podemos apreciar que la varianza de la serie incrementa cuando el nivel original de la serie incrementa, por lo que no es estacionaria. Superior derecha, la transformación logarítmica de la serie original. Inferior derecha, los rendimientos logarítmicos de la serie.

Aplicamos una prueba estadística conocida como la prueba Dickey-Fuller para conocer si los datos son o no estacionarios. Los resultados aplicados a las series originales nos demuestran que las series necesitan un tratamiento previo para ser estacionarias y poder así ajustar un modelo ARIMA. A continuación mostraremos las pruebas realizadas a los log-rendimientos de las series de acciones:

Para la serie de Bimbo:

Serie	Estadística Dickey-Fuller	Orden del rezago	P-value
BIMBO	-11.49	10	0.01
MEXCHEM	-9.94	10	0.01

Para ambas series de log-rendimientos no se rechaza que los datos sean estacionarios. Con estas pruebas hechas podemos comenzar el ajuste de los modelos ARIMA a las series de rendimientos. Es importante hacer dos señalamientos, las series originales les hemos realizado ya dos transformaciones, la primera fue aplicar logaritmos y la segunda fue la diferencia de esos logaritmos, por lo que debemos considerar esta primera diferencia en el valor del parámetro de diferencias  $d$  del modelo ARIMA.

Para determinar los valores de  $p$  y  $q$  en nuestros modelos, veamos cómo son el ACF y PACF de los datos:

Creamos una función en R que ajusta las posibles combinaciones de modelos con los valores

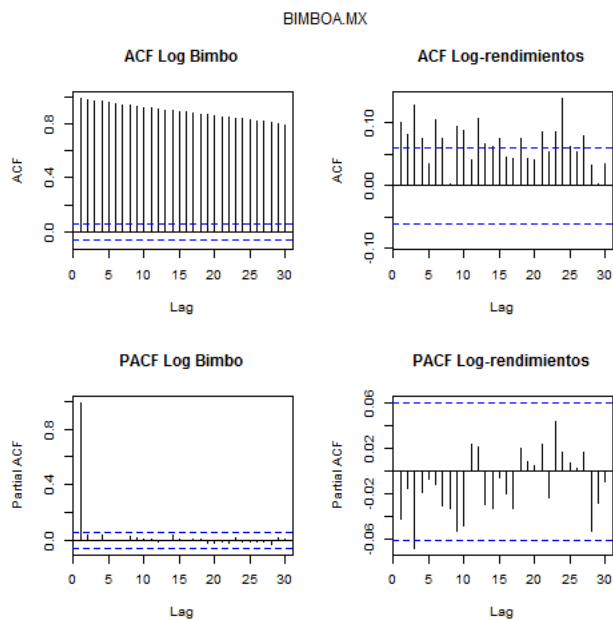


Figura 14: Acción BIMBO: Superior izquierda, ACF de los logaritmos de la serie. Inferior izquierda, PACF de los logaritmos de la serie. Superior derecha, ACF de los log-rendimientos de la serie. Inferior derecha, PACF de los log-rendimientos de la serie

$p, q, d \in \{0, 1, 2\}$ . Los modelos resultantes en ambas series son modelos ARIMA(1,0,1). El criterio utilizado fue el AIC. Procederemos a ajustar los modelos ARIMA a nuestros datos y analizar los residuales:

Ajuste de ARMA(1,1) a Bimbo:

	Coeficiente	Error estándar
AR	0.865	0.093
MA	-0.9101	0.0778

Medida	Valor
Log-verosimilitud	2848.35
AIC	-5690.7

Véase en el apéndice la tabla de comparación del AIC para varios modelos propuestos.

Y sus respectivos residuales son:

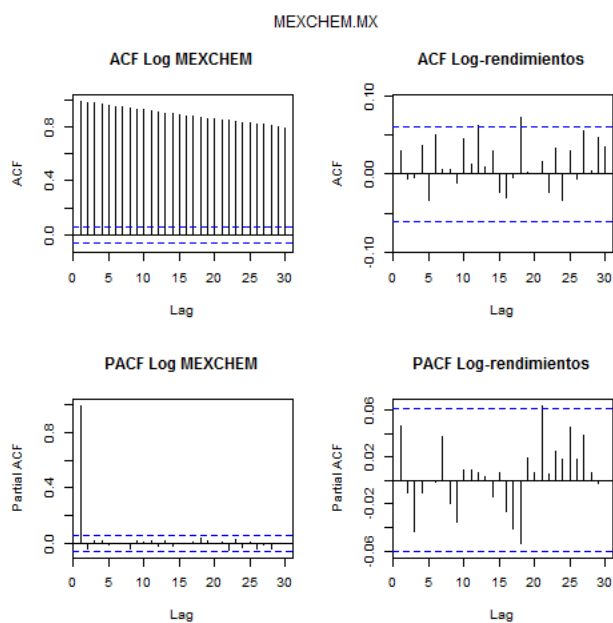


Figura 15: Acción MEXCHEM: Superior izquierda, ACF de los logaritmos de la serie. Inferior izquierda, PACF de los logaritmos de la serie. Superior derecha, ACF de los log-rendimientos de la serie. Inferior derecha, PACF de los log-rendimientos de la serie

Aunque podemos observar que el modelo con el menor AIC es un AR o un MA, al revisar el ACF y el PACF nos damos cuenta que el mejor modelo es una combinación de ambos, es decir un ARMA(1,1) aplicado a la serie de log-rendimientos. Los AIC's de otros modelos en la tabla nos indican que otros modelos pueden ser ajustados, sin embargo, elegiremos siempre ajustar el modelo que conlleve el menor número de parámetros.

Ajuste de ARMA(1,1) a Mexchem:

	Coefficiente	Error estándar
AR	0.0231	0.3914
MA	0.0240	0.3934

Medida	Valor
Log-verosimilitud	2831.33
AIC	-5654.67

Véase en el apéndice la tabla de comparación del AIC para varios modelos propuestos.

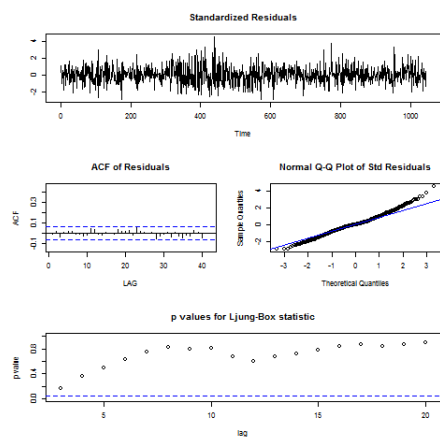


Figura 16: Superior: Residuales del modelo ARMA(1,1) ajustado a la serie de log-rendimientos BIMBO. Centro Izquierda: ACF de los residuales. Centro Derecha: Normal QQ-plot de los residuales estandarizados. Inferior: Representación gráfica de la prueba Ljung-Box aplicada a los primeros 20 rezagos.

Y sus respectivos residuales son:

Analicemos con detalle los residuales de los modelos ajustados a nuestras series:

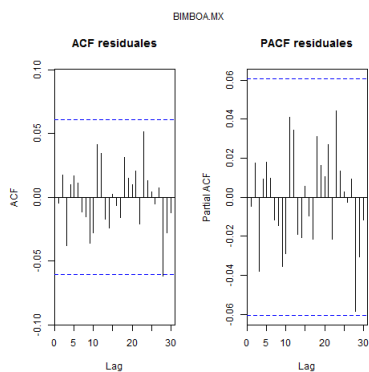


Figura 18: Izquierda: ACF de los residuales del modelo ajustado a los log-rendimientos de BIMBO. Derecha: PACF de los residuales del modelo ajustado a los log-rendimientos de BIMBO.



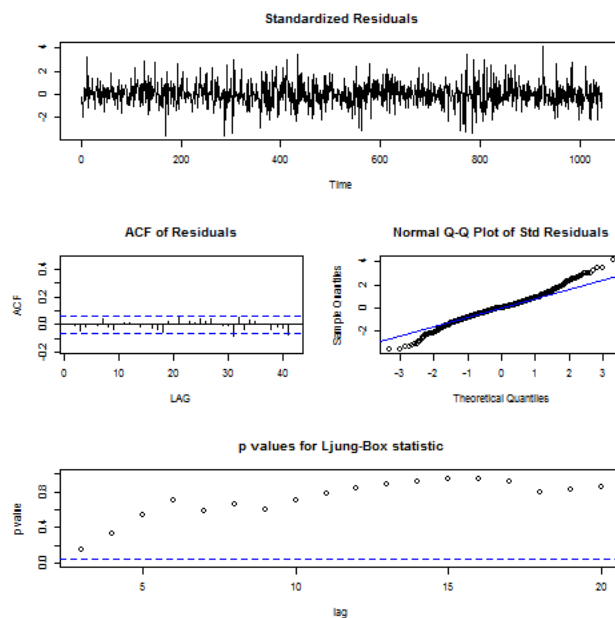


Figura 17: Superior: Residuales del modelo ARMA(1,1) ajustado a la serie de log-rendimientos MEXCHEM. Centro Izquierda: ACF de los residuales. Centro Derecha: Normal QQ-plot de los residuales estandarizados. Inferior: Representación gráfica de la prueba Ljung-Box aplicada a los primeros 20 rezagos.

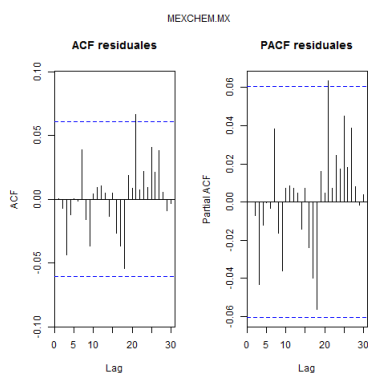


Figura 19: Izquierda: ACF de los residuales del modelo ajustado a los log-rendimientos de MEXCHEM. Derecha: PACF de los residuales del modelo ajustado a los log-rendimientos de MEXCHEM.

¿Se distribuyen nuestros residuales normales?

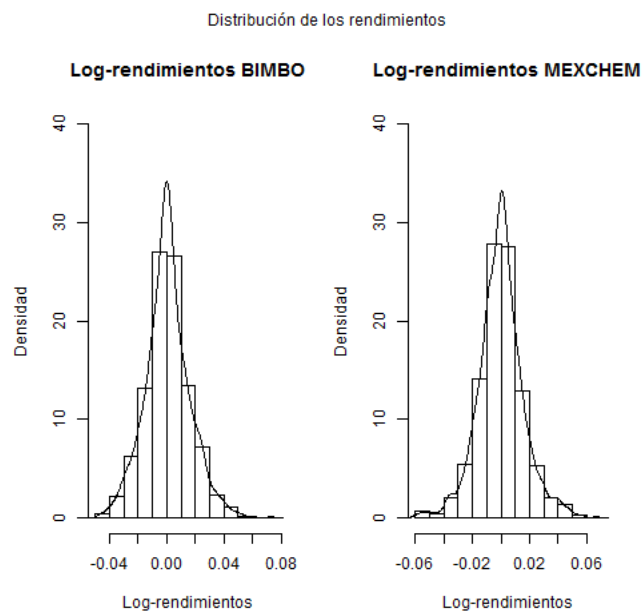


Figura 20: Izquierda: Distribución de los residuales del modelo ARMA(1,1) ajustado a la serie BIMBO. Derecha: Distribución de los residuales del modelo ARMA(1,1) ajustado a la serie MEXCHEM.

Revisemos si es necesario el ajuste de un modelo GARCH, para ello veamos cómo se ven los residuales al cuadrado.

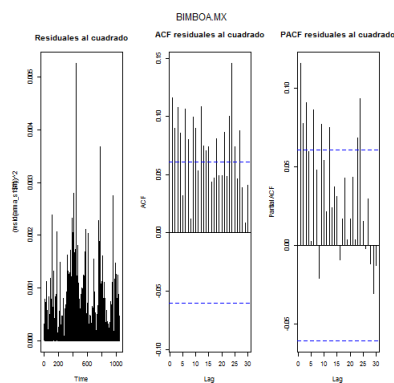


Figura 21: Izquierda: Residuales al cuadrado del modelo ARMA(1,1) ajustado a la serie BIMBO. Centro: ACF de los residuales al cuadrado. Derecha: PACF de los residuales al cuadrado.

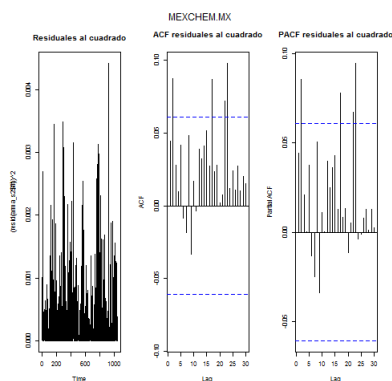


Figura 22: Izquierda: Residuales al cuadrado del modelo ARMA(1,1) ajustado a la serie MEX-CHEM. Centro: ACF de los residuales al cuadrado. Derecha: PACF de los residuales al cuadrado.

Sí necesitamos ajuste GARCH:

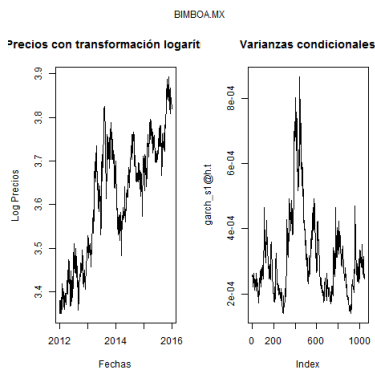


Figura 23: Izquierda: Serie de precios de BIMBO con transformación logarítmica. Derecha: Varianzas condicionales del modelo GARCH ajustado a los residuales del modelo ARMA(1,1).

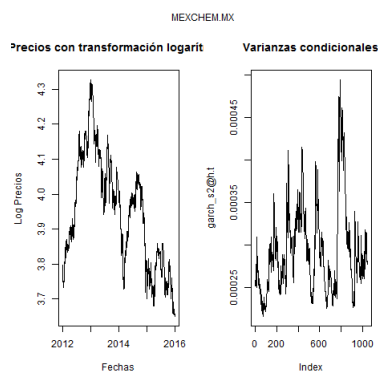


Figura 24: Izquierda: Serie de precios de MEXCHEM con transformación logarítmica. Derecha: Varianzas condicionales del modelo GARCH ajustado a los residuos del modelo ARMA(1,1).

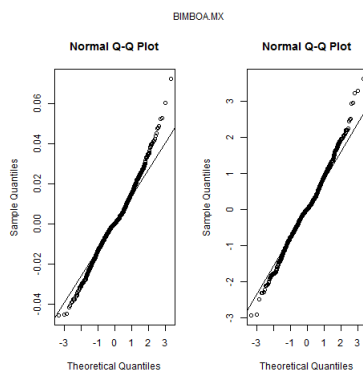


Figura 25: Izquierda: Q-Q plot de los residuos del modelo ARMA(1,1) ajustado a la serie BIMBO. Derecha: Q-Q plot de los residuos del modelo GARCH(1,1) ajustado a los residuos del modelo ARMA(1,1).

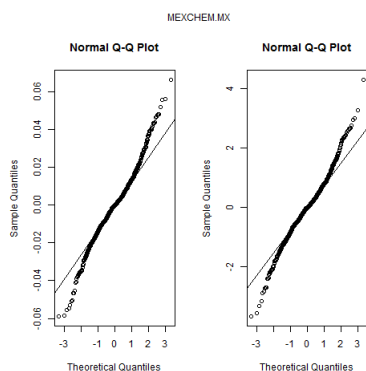


Figura 26: Izquierda: Q-Q plot de los residuales del modelo ARMA(1,1) ajustado a la serie MEXCHEM. Derecha: Q-Q plot de los residuales del modelo GARCH(1,1) ajustado a los residuales del modelo ARMA(1,1).

Las gráficas anteriores nos muestran que en ambos casos el efecto de las colas pesadas ha sido corregido por el modelo GARCH aunque no de manera íntegra.

A los residuales estandarizados de cada una de nuestras series, les ajustaremos una cópula. Antes de comenzar con el ajuste veamos como es la dependencia entre estas dos series:

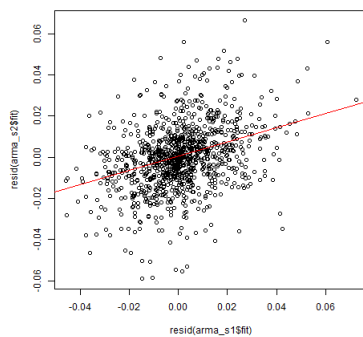


Figura 27: Gráfica de dispersión entre los residuales de los modelos ARMA(1,1)-GARCH(1,1) ajustados a ambas series de acciones.

## 9.6. Ajuste de la distribución Pareto Generalizada y cópula

Ajustaremos a cada una de las series de residuales estandarizados, la distribución Pareto Generalizada. Graficaremos la distribución empírica original para los datos de acción de BIMBO:

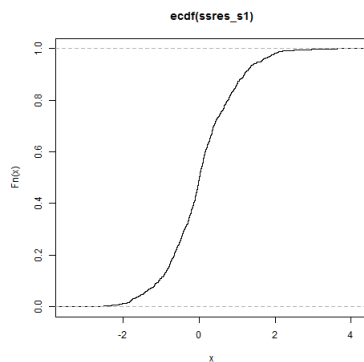


Figura 28: Función de distribución empírica de los residuales del modelo GARCH(1,1) ajustado a los residuales del modelo ARMA(1,1) a los log-rendimientos de BIMBO.

y para los de la acción de MEXCHEM:

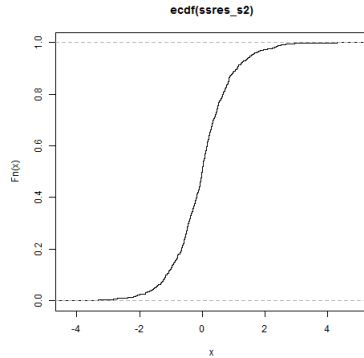


Figura 29: Función de distribución empírica de los residuales del modelo GARCH(1,1) ajustado a los residuales del modelo ARMA(1,1) a los log-rendimientos de MEXCHEM.

Y este es el ajuste de una distribución Pareto Generalizada a cada serie. El motivo de este ajuste es que, como podemos apreciar, en las gráficas superiores la cantidad de observaciones en la cola de la distribución es muy pequeña pero que son los casos extremos que nos son de suma importancia, por lo que el ajuste de esta distribución robustecerá el ajuste de las colas:

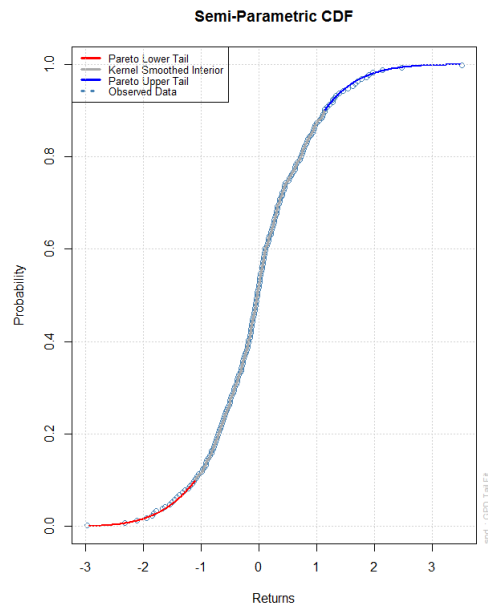


Figura 30: Función de distribución de la Pareto Generalizada ajustada a los residuales del modelo ARMA(1,1)-GARCH(1,1).

Cuando evaluamos nuestros residuales estandarizados en cada una de las funciones de distri-

bución Pareto Generalizada ajustadas, obtenemos valores entre  $[0, 1]$ . A continuación, analizaremos la estructura de dependencia de esta transformación de residuales:

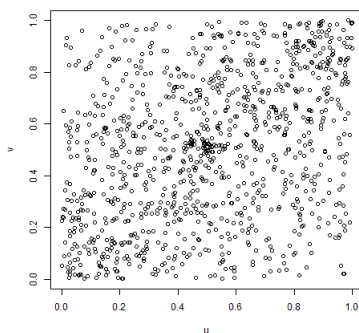


Figura 31: Gráfica de dispersión entre las marginales de la cópula.

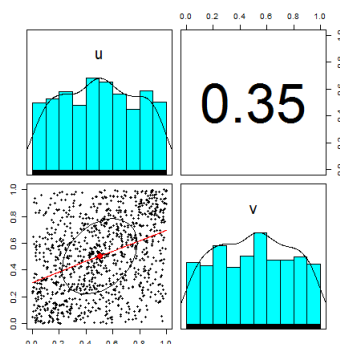


Figura 32: Superior izquierda: Histograma de la marginal de la serie BIMBO. Inferior izquierda: Gráfica de dispersión entre las marginales de la cópula. Superior derecha: Correlación lineal entre las marginales de la cópula. Inferior derecha: Histograma de la marginal de la serie MEXCHEM.

El valor de cada una de las medidas de asociación es el siguiente:

Medida	Valor
Pearson	0.3471863
Kendall	0.2390473
Spearman	0.3463974



Hemos decidido ajustar una cópula t-Student<sup>12</sup> a los datos. Dicha cópula necesita dos parámetros: los grados de libertad y de la correlación. En la literatura, hemos encontrado de manera recurrente el uso de 5 grados de libertad, ya que si el valor crece, entonces se estaría ajustando una cópula normal<sup>13</sup>. Dicho lo anterior, sólo resta conocer el valor del segundo parámetro.

Para conocer el valor del parámetro, utilizaremos la estadística Von Mises, descrita líneas arriba. Lo que haremos será comparar la cópula empírica con diversas cópulas t, en las que el parámetro vaya cambiando su valor y nos quedaremos con aquel valor que minimice la distancia entre estas cópulas. Hemos programado una función en R para este fin:

$\rho$	Estadística
0.1	0.8004137
0.2	0.3424068
0.3	0.07962807
0.37	0.01758788
0.4	0.02333812
0.5	0.1926576

Nos interesa el parámetro que haga que nuestra estadística tenga el menor valor. como podemos ver en el cuadro anterior, el parámetro que buscamos se encuentra en el intervalo:  $[\cdot, 3, \cdot 4]$ . Refinando los cálculos, obtenemos un valor final de 0.3701145.

La densidad de la cópula ajustada es:

En esta tesis se propone otro método para el ajuste de la cópula, utilizando el estimador de la función de dependencia extrema. Seguiremos el siguiente procedimiento:

1. Ajustaremos el estimador de la función de dependencia utilizando nuestros datos y considerando  $\frac{k}{n} = \frac{100 \cdot \log(n)}{n}$ .<sup>14</sup>
2. Se calcula un estimador del coeficiente de dependencia para la cópula t, que se encuentra dado en términos de los grados de libertad y de  $\rho$  que son los parámetros de la cópula. En este punto, fijaremos uno de los parámetros, los grados de libertad, y buscaremos el valor de  $\rho$  que minimice la distancia entre el estimador del coeficiente de dependencia y el estimador de la función de dependencia.

El coeficiente de dependencia de cola (superior e inferior) de una cópula t-Student se encuentra dado por:

<sup>12</sup>Se propone utilizar este tipo de cópula ya que ambos rendimientos presentan una curtosis alta

<sup>13</sup>Recordemos que una cópula es una función de distribución con marginales uniformes.

<sup>14</sup>Con n igual al número de datos de cada serie.

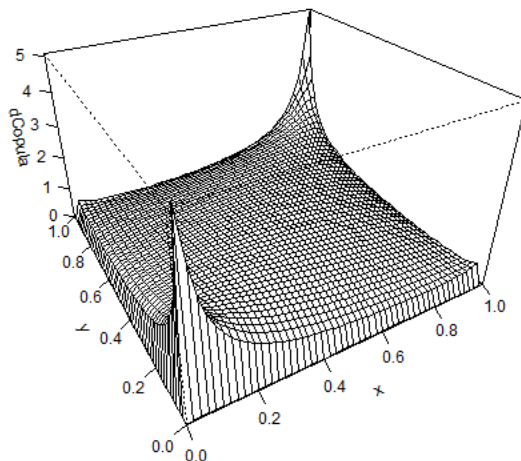


Figura 33: Densidad de la cópula ajustada.

$$\begin{aligned}
 \Lambda_{L,\rho}^t(u,v) &= \Lambda_{U,\rho}^t(u,v) \\
 &= u * t_{\nu+1}\left(-\sqrt{\frac{\nu+1}{1-\rho^2}}\left(\left(\frac{u}{v}\right)^{\frac{1}{\nu}} - \rho\right)\right) \\
 &\quad + v * t_{\nu+1}\left(-\sqrt{\frac{\nu+1}{1-\rho^2}}\left(\left(\frac{v}{u}\right)^{\frac{1}{\nu}} - \rho\right)\right)
 \end{aligned}$$

con  $t_{\nu+1}$  la función de distribución t-Student estándar con  $\nu + 1$  grados de libertad.

Bajo este enfoque, se hicieron algunas pruebas con diferentes valores para  $\rho$ , se obtuvo lo siguiente:

$\rho$	Estadística
0.1	13.5318
0.2	11.68887
0.7	0.8886351
0.8	0.8537409
0.81	1.011759
0.9	5.020115

Es decir, bajo este método, lo mejor sería ajustar un modelo t-Student de parámetros  $\rho = ,8$  y  $\nu = 5$ .

La siguiente gráfica muestra una comparación de ambos métodos:

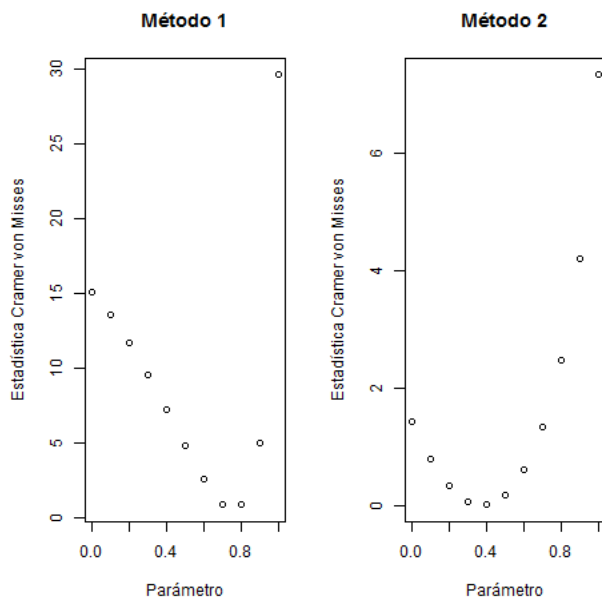


Figura 34: Izquierda: Estadística Crámer Von Misses para el método 1. Derecha: Estadística Crámer Von Misses para el método 2.

Calculamos también el valor de los coeficientes de cola superior e inferior y obtuvimos el siguiente valor: 0.1478116 <sup>15</sup>

## 9.7. Simulaciones finales

Se realizaron 100 simulaciones del modelo presentado en la sección anterior. Las siguientes gráficas muestran las simulaciones de los precios para 21 días hábiles (los días hábiles de enero del 2016). Debido a que se conocían los datos que realmente pasaron, se muestran en las gráficas correspondientes, y en general, podemos ver que alguna de las trayectorias simuladas tienen el valor del dato real.

Las siguientes gráficas muestran las simulaciones de precios para ambas series.

<sup>15</sup>Recordar que ambos coeficientes son los mismos en el caso de la cópula t.

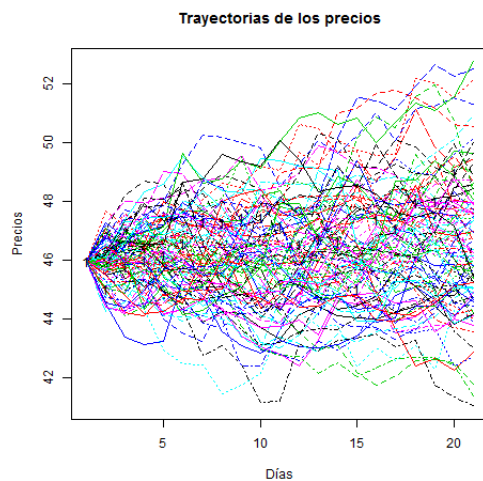


Figura 35: 100 trayectorias simuladas para 20 días hábiles del precio de BIMBO. La trayectoria real del precio se muestra con marcadores.

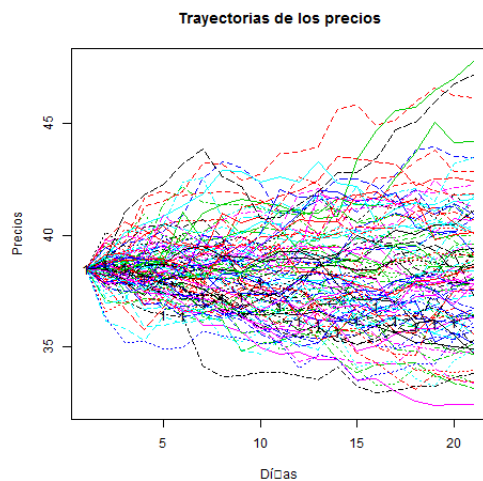


Figura 36: 100 trayectorias simuladas para 20 días hábiles del precio de MEXCHEM. La trayectoria real del precio se muestra con marcadores.

### 9.8. Cálculo del VaR

Suponiendo un portafolio con pesos iguales para cada activo podemos conocer el comportamiento del VaR para las simulaciones:

Máxima pérdida simulada: 14.46 Máxima ganancia simulada: 12.26

90 %	95 %	91 %
5.98	7.77	6.32

## 9.9. Resumen del capítulo

- La creciente necesidad de una acertada medición de los riesgos que involucran las operaciones financieras ha promovido el desarrollo de ciertos modelos y medidas como lo son las cópulas y el valor en riesgo (VaR).
- En la aplicación nos percatamos de que el ajuste de los modelos de series de tiempo ARMA-GARCH, las distribuciones marginales, así como de la cópula juegan un papel crucial para que el modelo funcione de manera adecuada.
- Con el procedimiento aplicado, nos dimos cuenta de que el modelo ajustado es adecuado para nuestros datos ya que el mes simulado cubre el escenario de cada una de las series quedando en la media.

## 10. Conclusiones generales

- Las cópulas son una herramienta muy útil en el modelado de series financieras ya que ofrecen gran flexibilidad para la estructura de dependencia. Conociendo la función de distribución conjunta podemos conocer las funciones de distribución marginales. Sin embargo, el conocer las funciones de distribución marginales no nos permite construir a la función de distribución conjunta. Aún nos hace falta conocer la manera en la que se inter-relacionan éstas funciones de distribución marginales, la cópula es la función que nos hace falta y que cumple con esta función.
- Hoeffding y Fréchet determinaron que toda cópula recae entre ciertas cotas, debido a la existencia de casos de dependencia extrema. Las cópulas fundamentales son tres casos de dependencia extrema: independencia, dependencia positiva perfecta (co-monotonidad) y dependencia negativa perfecta (contra-monotonidad). La cópula co-monotónica existe para toda dimensión  $d \geq 2$  pero la contra-monotona sólo para el caso bidimensional.
- Analizamos dos clasificaciones de cópulas: implícitas y explícitas. De las primeras se pueden extraer estructuras de dependencia bien conocidas. Entre ellas, incluyen pero no se limitan a las cópulas elípticas, asociadas a las distribuciones elípticas.
- Las cópulas explícitas, por el contrario, tienen formas sencillas y cerradas. En esta categoría se encuentran las cópulas arquimedianas (se crean utilizando la función generadora de cópulas), por ejemplo: las cópulas Clayton, Gumbel, Frank, entre otras. En este grupo también se encuentra la cópula de Marshall Olkin.

- La correlación lineal tiene algunas debilidades para ser tomada como medida de dependencia, entre ellas se encuentra el hecho de que es invariante únicamente ante transformaciones lineales (no bajo cualquier tipo de transformación).
- Además, la correlación lineal es importante especialmente en el caso de distribuciones elípticas y está definido únicamente cuando la varianza de ambas variables es finita, lo cual puede representar un obstáculo cuando se trabaja con distribuciones de cola pesada.
- Las tau de Kendall y la rho de Spearman son medidas de asociación que pueden ser expresadas en términos de cópulas, por eso suelen ser una alternativa a la correlación lineal. Aunque dan como resultado un número y no dan mayor descripción de la estructura de dependencia.
- La dependencia de cola es una medida utilizada para describir la magnitud de los valores que aparecen con variaciones en la otra variable. En la tesis abordamos coeficientes y funciones de dependencia extrema. Las funciones y coeficientes de dependencia extrema pueden escribirse en términos de cópulas.
- En 1979, Deheuvels enunció una forma de estimar a la cópula por la hoy llamada cópula empírica. Recordando que la cópula es una función de distribución, es lógico pensar en la función de distribución empírica como estimador para la cópula. Además, Deheuvels demostró que cuando el número de observaciones crece, la cópula empírica tiende a la cópula verdadera.
- Dependiendo en la información disponible así como de la finalidad del estudio, se pueden estimar los parámetros de la cópula por máxima verosimilitud completa, paso a paso o semi-paramétrica. Una de las ventajas que tiene la estimación semi-paramétrica es que no se necesita ningún supuesto para la distribución de las marginales, pues estas son sustituidas por sus correspondientes funciones de distribución empírica.
- Una vez que hemos ajustado los parámetros de la cópula es necesario conocer si la cópula seleccionada es la que mejor ajusta a nuestros datos, entre las pruebas que existen para probar la bondad de ajuste de la cópula está la estadística Crámer von Misses que calcula la suma de los cuadrados de las distancias entre la cópula empírica y la cópula elegida.
- La función de dependencia extrema es la generalización de los coeficientes de dependencia extrema expresada como un límite. Por su definición es difícil encontrar un estimador adecuado y mucho se ha escrito al respecto. Aquí se aborda como un estimador a la cópula empírica evaluada en puntos que cumplen con características asintóticas.
- Para la simulación de cópulas se procede de manera contraria que con la estimación, ya que primero se simula la dependencia y luego las marginales. Lo anterior se debe a

que al utilizar las cópulas se puede modelar de manera independiente la estructura de dependencia de las marginales.

- La creciente necesidad de una acertada medición de los riesgos que involucran las operaciones financieras ha promovido el desarrollo de ciertos modelos y medidas como lo son las cópulas y el valor en riesgo (VaR).
- En la aplicación nos percatamos de que el ajuste de los modelos de series de tiempo ARMA-GARCH, las distribuciones marginales, así como de la cópula juegan un papel crucial para que el modelo funciones de manera adecuada.
- Con el procedimiento aplicado, nos dimos cuenta de que el modelo ajustado es adecuado para nuestros datos ya que el mes simulado cubre el escenario de cada una de las series quedando en la media.
- En la realización de la parte práctica de esta tesis me percaté sobre la importancia de un buen ajuste a las marginales. El uso de cópulas para el cálculo del VaR resulta de gran utilidad ya que permite flexibilidad para modelar a las marginales de manera independiente a la estructura de dependencia. En esta tesis específicamente, ajustamos distribuciones Pareto Generalizadas con la finalidad de robustecer el modelado marginal debido a la pocas observaciones que tenemos en las colas.
- Una aportación poca explorada que se propone en esta tesis es la determinación del parámetro de la cópula utilizando la función de dependencia de cola, lo cual trae como beneficio considerar el comportamiento de la cópula en las colas.
- Se podría profundizar en el tema aumentando el número de activos que posee el portafolio (en nuestro caso, realizamos los cálculos de un portafolio de dos activos (situación poco común en la práctica). Una alternativa a nuestro enfoque podría ser utilizar la teoría de valores extremos.

## 11. Apéndice

Comparación de los AIC de los modelos ajustados a Bimbo:

Modelo	P	D	Q	AIC
1	0	0	0	-7.290587
2	0	0	1	-7.290480
3	0	0	2	-7.288957
4	0	1	0	-6.556274
5	0	1	1	-7.287747
6	0	1	2	-7.287637
7	0	2	0	-5.450414
8	0	2	1	-6.555314
9	0	2	2	-7.264367
10	1	0	0	-7.290417
11	1	0	1	-7.296250
12	1	0	2	-7.294465
13	1	1	0	-6.860757
14	1	1	1	-7.287578
15	1	1	2	-7.291100
16	1	2	0	-6.075927
17	1	2	1	-6.859795
18	1	2	2	-7.280175
19	2	0	0	-7.288733
20	2	0	1	-7.294478
21	2	0	2	-7.293315
22	2	1	0	-6.956248
23	2	1	1	-7.285894
24	2	1	2	-7.284017
25	2	2	0	-6.329413
26	2	2	1	-6.955283
27	2	2	2	-7.278723



Comparación de los AIC de los modelos ajustados a Mexchem:

Modelo	P	D	Q	AIC
1	0	0	0	-7.262964
2	0	0	1	-7.263269
3	0	0	2	-7.261369
4	0	1	0	-6.616795
5	0	1	1	-7.262253
6	0	1	2	-7.262341
7	0	2	0	-5.538825
8	0	2	1	-6.615834
9	0	2	2	-7.240469
10	1	0	0	-7.263238
11	1	0	1	-7.26134
12	1	0	2	-7.260256
13	1	1	0	-6.866194
14	1	1	1	-7.262324
15	1	1	2	-7.260287
16	1	2	0	-6.101439
17	1	2	1	-6.865233
18	1	2	2	-7.241812
19	2	0	0	-7.261427
20	2	0	1	-7.260416
21	2	0	2	-7.26029
22	2	1	0	-6.95752
23	2	1	1	-7.260567
24	2	1	2	-7.258579
25	2	2	0	-6.346962
26	2	2	1	-6.956558
27	2	2	2	-7.244884

# Bibliografía

- [1] Carmona, René. (2014). *Statistical Analysis of Financial Data in R*. Nueva York: Springer.
- [2] Cont, Rama. (2001). *Empirical properties of asset returns: stylized facts and statistical issues* Institute of Physics Publishing, Quantitative Finance Vol.1, Páginas 223-236.
- [3] Joe, Harry. (2014). *Dependence Modelling with Copulas*. Nueva York: Springer.
- [4] McNeil, A. J., Frey R. y Embrechts P. (2005). *Quantitative Risk Management: Concepts, Techniques and Tools*. Princeton: Princeton University Press.
- [5] Nelsen, R.B.(2006). *An introduction to Copulas*. Nueva York: Springer.
- [6] Salazar, Yuri.(2010). *Negative (LU and UL) tail dependence using copulae* University of Essex.
- [7] Schmidt, R. y Stadtmüller, U. (2006). *Non-Parametric Estimation of Tail Dependence*. Scandinavian Journal of Statistics, Vol. 33 (Tema 2), Páginas 307–335.
- [8] Schmidt, T. (Diciembre 2006). *Coping with copulas*. En *Copulas from theory to application in finance*(Sección 1, Capítulo 1). Londres: Risk Books.
- [9] Zhang, M. -H.,(2007) *Modelling total tail dependence along diagonals* Insurance: Mathematics and Economics Vol. 42, Páginas 77-80.