

# Construcción de Distribuciones Multivariadas con Marginales Dependientes Usando Cópulas en R

## Construction of Multivariate Distributions with Dependent Marginals Using Copulas in R

M. C. Jaramillo-Elorza<sup>a\*</sup>

J. A. Lozano<sup>b,\*</sup>

Recepción: 03-mar-13

Aceptación: 27-mar-13

### Resumen

Las cópulas se han convertido en una herramienta popular para la construcción de modelos multivariados en campos donde la dependencia multivariada es de gran interés. El propósito de este trabajo es presentar las cópulas tanto en su concepto teórico, como en su implementación en el software estadístico R y profundizar en la construcción de distribuciones multivariadas con marginales dependientes, usando la clase *mvdc* del paquete *copula*, la cual permite utilizar varias y diferentes marginales ya implementadas. Además, se trabajará con métodos para dibujar representaciones de perspectiva y contorno para las funciones de distribución y densidad.

**Palabras clave:** Análisis Multivariado, Cópulas en R, Paquete copula, Software R.

### Abstract

The copula has become a popular tool to build the multivariate models, in many fields where the multivariate dependence is of a great interest. This paper purpose is to present the copula both in their theoretical concept and its implementation in the R statistical software, and to deepen into the multivariate distributions? construction with the dependent marginal, by using the *copula* package's *mvdc* class, which allows to use the marginal in several and different types, that have been implemented already. In addition, to work with the methods for drawing the perspective and the contour representations for the distribution and the density functions.

**Key words:** Copula, Copula package, Multivariate Analysis, R Software.

---

<sup>a</sup>Profesor Asociado, Escuela de Estadística, Universidad Nacional de Colombia-Medellín.

<sup>b</sup>Estudiante de Maestría en Estadística, Universidad Nacional de Colombia-Medellín.

\*Correo electrónico: mcjarami@unal.edu.co

## 1. Introducción

Las cópulas se han convertido en una herramienta popular para la construcción de modelos multivariados en campos donde hay un gran interés por la dependencia multivariada; más específicamente, han sido utilizadas en campos tan variados como en medicina, para modelar el número de casos mensuales de una enfermedad casi erradicada, por ejemplo, la poliomielitis [1]; en ciencias actuariales, para modelar la mortalidad y las pérdidas dependientes [2, 3]; en finanzas, para análisis y gestión de riesgos [4, 5]; en estudios biomédicos, en el modelado de tiempos de eventos correlacionados y riesgos competitivos [6], y en ingeniería, en el control de procesos multivariados y en el modelado hidrológico [7].

En [8] se describen las cópulas como “... funciones de distribución multivariada que unen o ‘copulan’ sus funciones de distribución marginales unidimensionales”, y como “... funciones de distribución multivariada cuyas marginales unidimensionales son uniformes en el intervalo  $[0,1]$ ”. Las cópulas son muy útiles a la hora de derivar distribuciones conjuntas, conociendo las distribuciones marginales, especialmente cuando las variables son no normales; por otro lado, se utilizan en un contexto bivariado para definir medidas no paramétricas de dependencia [9].

Una idea más formal de lo que es una cópula se da al considerar un par de variables aleatorias  $X$  y  $Y$  con funciones de distribución marginales  $F(x) = P[X \leq x]$  y  $G(y) = P[Y \leq y]$ , respectivamente, y función de distribución conjunta  $H(x,y) = P[X \leq x, Y \leq y]$ . Ahora bien, al par de números  $(x,y)$  se le puede asociar tres números  $F(x)$ ,  $G(y)$  y  $H(x,y)$ , donde cada uno de ellos pertenece al intervalo  $[0,1]$ , es decir, a cada  $(x,y)$  le corresponde un punto  $(F(x), G(y))$  en el espacio producto  $[0, 1] \times [0, 1]$ , y a este par ordenado le corresponde un número  $H(x,y)$  en  $[0,1]$  (ver [8]).

$$C : [0, 1] \times [0, 1] \rightarrow [0, 1]$$

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

Como se demuestra en [8], esta correspondencia, que asigna el valor de la función de distribución conjunta a cada par ordenado de los valores de las funciones de distribución marginales, es de hecho una función de distribución. Tales funciones son cópulas [10]. Más formalmente, en [11] se describe una cópula como una distribución multivariada cuyas

marginales son uniformes en  $[0,1]$ . Para un vector  $p$ -dimensional sobre el cubo unitario, una cópula es

$$C(u_1, \dots, u_p) = P(U_1 \leq u_1, \dots, U_p \leq u_p)$$

Por otro lado, [12] afirman que el teorema subyacente de las cópulas fue introducido por Sklar, primero en 1959, en un artículo en francés, y luego, en 1973, en un artículo similar en inglés; hay que tener en cuenta que aunque el término cópula fue introducido por Sklar (1959), ideas y resultados similares pueden ser rastreados en varios trabajos, más notablemente en Hoeffding (1940, 1941). Las posibles mejores bandas para estas funciones las estableció [13], quien estudió medidas de dependencia que son invariantes bajo transformaciones estrictamente crecientes.

## 2. Teorema de Sklar

Para introducir el concepto del teorema de Sklar se hará referencia, inicialmente, al caso bivariado. Sean  $X$  y  $Y$  variables aleatorias continuas con funciones de distribución  $F$  y  $G$ , respectivamente, y sea  $H$  su función de distribución conjunta; entonces, existe una cópula tal que para todo  $x,y$ , reales se tiene

$$H(x,y) = C(F(x), G(y)) \quad (1)$$

Si  $F$  y  $G$  son continuas, entonces  $C$  es única [8]. Desde otra perspectiva, esto puede ser interpretado de manera inversa, es decir, si  $C$  es una cópula, y  $F$  y  $G$  son funciones de distribución univariadas, entonces es posible determinar la función de distribución conjunta  $H$  dada por (1.1), con marginales  $F$  y  $G$ .

En general, una función de distribución  $n$ -variada puede ser determinada con la versión  $n$ -variada del Teorema de Sklar:

Sean  $X_1, X_2, \dots, X_n$  variables aleatorias continuas con funciones de distribución  $F_1, F_2, \dots, F_n$ , respectivamente, y función de distribución  $n$ -dimensional conjunta  $H$ . Entonces, existe una  $n$ -cópula  $C$  tal que para todo vector  $X$  real se tiene

$$H(x_1, x_2, \dots, x_n) =$$

$$C(F_1(x_1), F_2(x_2), \dots, F_n(x_n)) \quad (2)$$

Si  $F_1, F_2, \dots, F_n$  son todas continuas, entonces es única (ver [9]).

Además, como se menciona en el título de este artículo, las marginales deben ser dependientes para que el uso de cópulas en la construcción de distribuciones multivariadas sea útil, ya que si son independientes la distribución conjunta se reduciría solamente al producto de las marginales univariadas.

Para hacer mención de esta dependencia de las marginales se puede hacer referencia a [9], quien, considerando que  $C_\alpha$  es una función de distribución con densidad  $c_\alpha$  sobre  $[0, 1] \times [0, 1]$  para  $\alpha$  en  $\mathbb{R}$  y denotando  $T_1, T_2$  los tiempos de falla, y  $(S_2, S_2)$ ,  $(f_1, f_2)$  las respectivas funciones de confiabilidad y de densidad marginales, describe así, a través del modelo cópula, las funciones de confiabilidad y densidad conjuntas de  $T_1$  y  $T_2$ :

$$S(t_2, t_2) = C_\alpha(S_1(t_1), S_2(t_2)), \quad t_1, t_2 \geq 0$$

$$f(t_2, t_2) = C_\alpha(S_1(t_1), S_2(t_2))f_1(t_1)f_2(t_2), \quad t_1, t_2 \geq 0$$

donde  $(T_1, T_2)$  proviene de una cópula  $C_\alpha$  (ver [8]).

Este valor de  $\alpha$  es, en realidad, el que une o copula las distribuciones, y está asociado al parámetro de dependencia entre los tiempos de falla  $T_1$  y  $T_2$ , el cual puede estar relacionado con la correlación o con alguna clase de matriz de correlación para casos multivariados o, entre otros, con el  $\tau$  de Kendal, que va a estar completamente determinado por la cópula [14].

Dos de las familias de cópulas más usadas son: las cópulas Elípticas y las Arquimedianas.

### 3. Familia Elíptica

Esta familia de cópulas se caracteriza por compartir propiedades de la normal multivariada [13], tales como simetría, o el hecho de que la estructura de dependencia esté totalmente determinada por la matriz de correlación; facilitan la obtención de modelos multivariantes para extremos y otras formas de dependencia no normales, y son de gran importancia en finanzas y manejo de riesgos, por su fácil implementación. Se definen como las cópulas asociadas a las distribuciones elípticas por el teorema de Sklar, como, por ejemplo, las cópulas Normales y t-Student. Sean  $F$  una c.d.f. de una distribución elíptica multivariada y  $F_i$  la c.d.f. de  $i$ -ésima marginal, y  $F_i^{-1}$  su respectiva función inversa,  $i = 1, \dots, n$ , la cópula elíptica determinada por  $F$  es

$$C(u_1, u_2, \dots, u_n) = F(F_1^{-1}(u_1), F_2^{-1}(u_2), \dots, F_n^{-1}(u_n))$$

Se llama cópula elíptica bivariada a toda cópula de la forma:

$$C_\rho(u, v) = \frac{1}{\sqrt{1-\rho^2}} \times \int_{-\infty}^{\Phi_{g,1}^{-1}(u)} \int_{-\infty}^{\Phi_{g,2}^{-1}(v)} g\left(\frac{x^2 - 2\rho xy + y^2}{\sqrt{1-\rho^2}}\right) dx dy$$

$$= H_\rho(\Phi_{g,1}^{-1}(u), \Phi_{g,2}^{-1}(v))$$

La cual es la distribución conjunta de las variables  $X$  y  $Y$ ,  $\Phi_{g,1}^{-1}(u)$ ,  $\Phi_{g,2}^{-1}(v)$ , son las respectivas funciones cantiles, y  $\rho$  es el coeficiente de correlación entre  $X$  y  $Y$ .

La cópula normal bivariada tiene la forma:

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

$$= \int_{-\infty}^{\Phi^{-1}(u)} \int_{-\infty}^{\Phi^{-1}(v)} \frac{1}{2\pi\sqrt{1-\theta^2}} \times \exp\left(\frac{-(x^2 - 2\theta xy + y^2)}{2(1-\theta^2)}\right) dx dy$$

donde  $\Phi$  es la c.d.f. de la distribución normal estándar, y  $\phi_G(u, v)$  es la distribución normal estándar bivariada con parámetro de correlación  $\theta$  en  $(-1, 1)$  [15].

Por otro lado, un ejemplo de una cópula elíptica con dos parámetros de dependencia es la distribución  $t$  bivariada con  $w$  grados de libertad y correlación  $\rho$ :

$$C(u, v) = \int_{-\infty}^{t_{\theta_1}^{-1}(u)} \int_{-\infty}^{t_{\theta_2}^{-1}(v)} \frac{1}{2\pi\sqrt{1-\theta_2^2}} \times \left(1 + \frac{x^2 - 2\theta_2 xy + y^2}{w(1-\theta_2^2)}\right) dx dy$$

donde  $t_{\theta_1}^{-1}(u)$  denota la inversa de la c.d.f de la distribución  $t$  univariada estándar con  $\theta_1$  grados de libertad, y  $t_{\theta_2}^{-1}(v)$  denota la inversa de la c.d.f de la distribución  $t$  univariada estándar con  $\theta_2$  grados de libertad. Los dos parámetros de dependencia son  $(\theta_1, \theta_2)$ ;  $\theta_1$  controla la pesadez de las colas [15].

### 4. Familia Arquimediana

Las cópulas Arquimedianas no surgieron originalmente en la estadística, aparecieron en el estudio de espacios métricos probabilísticos, donde fueron estudiadas como parte del desarrollo de una versión

probabilística de la desigualdad triangular. La selección de esta familia se debe a que se puede simular fácilmente; además, los cálculos de medidas de dependencia se simplifican, lo cual permite una mejor estimación de los parámetros. La distribución multivariada de esta familia para  $p$  variables está representada por:

$$C(u_1, \dots, u_p) = \phi^{-1} \left[ \sum_{i=1}^p \phi(u_i) \right]$$

donde  $\phi$  es el generador de la cópula  $C$ . Para el caso de cópulas bivariadas, la familia Arquimediana se representa por (ver [16]):

$$C_\alpha(u, v) = \phi_\alpha^{-1} [\phi_\alpha(u) + \phi_\alpha(v)], 0 \leq u, v \leq 1$$

donde  $\phi_\alpha$  se denomina el generador de la cópula  $C_\alpha$ , el cual es una función convexa y decreciente tal que  $\phi_\alpha > 0$ . Además, su inversa  $\phi_\alpha^{-1}$  es la transformada de Laplace de una variable latente denotada  $\tau$ , la cual induce la dependencia  $\alpha$ . La tabla 1 muestra las cópulas que se considerarán de la familia Arquimediana y su espacio parametral:

**Tabla 1.** Cópulas arquimedianas bivariadas y su respectivo espacio parametral.

Familia cópula	Espacio parametral	Cópula bivariada $C_\alpha(u, v)$
Clayton	$\alpha > 0$	$[u^{-\alpha} + v^{-\alpha} - 1]^{-1/\alpha}$
Gumbel	$\alpha \geq 1$	$\exp\{-[(-\ln u)^\alpha + (-\ln v)^\alpha]^{-1/\alpha}\}$
Frank	$-\infty < \alpha < \infty$	$\frac{-1}{\alpha} \ln[1 + (e^{-\alpha u} - 1)(e^{-\alpha v} - 1)/(e^{-\alpha} - 1)]$

Además, es necesario tener en cuenta los generadores y el coeficiente de correlación  $\tau$  de Kendall:

**Tabla 2.** Generadores de cópulas arquimedianas bivariadas y su respectivo de Kendall.

Familia Cópula	Generador	$\tau$ de Kendall
Clayton	$t^{-\alpha} - 1$	$(\alpha)/(\alpha + 2)$
Gumbel	$(-\ln t)^\alpha$	$\frac{\alpha-1}{\alpha}$
Frank	$\ln \frac{e^{-\alpha t} - 1}{e^{-\alpha} - 1}$	$1 - \frac{4}{\alpha} \left(1 - \frac{1}{\alpha} \int_0^\alpha \frac{t}{e^t - 1} dt\right)$

## 5. Experimental

### 5.1. Implementación al software R

Las técnicas existentes para estimar distribuciones conjuntas de datos no lineales son a menudo tarea difícil, ya que requieren exigentes procedimientos de estimación computacional basados en simulaciones.

Aunque los fundamentos de las cópulas son complejos o, como se afirma en [17], todavía no son muy fáciles de usar, su implementación práctica y simulación ahora es relativamente sencilla [12].

La implementación de cópulas en paquetes estadísticos es muy escasa en cuanto a variedad, aunque no en desarrollo; por ejemplo, el software comercial *Splus* cuenta con el módulo *finmetrics*, que genera números aleatorios de las funciones de densidad y distribución de una cópula bivariada, y las ajusta a unos datos [11]. Dado que este software tiene muchas limitaciones y es comercial, se dificulta el desarrollo de métodos y aplicaciones para cópulas, para lo cual es necesaria una plataforma de uso libre, como lo es R, que funciona en muchas plataformas, genera muy buenas gráficas, ofrece ayuda en línea y cuenta con muchas funcionalidades, como funciones de densidad y distribución multivariadas de la distribución normal y  $t$ , las cuales pueden ser usadas para cópulas.

Las ventajas de R llevaron a Yan a diseñar el paquete *copula*, usando características orientadas a objetos del lenguaje S y del estilo S4 para *cópulas* elípticas y Arquimedianas. El paquete *copula* de [11] cuenta con dos clases: *copula* y *mvdc*. Para utilizar los códigos que se usarán a continuación, es necesario cargar los paquetes:

```
library(pspline)
library(mvtnorm)
library(sn)
library(mnormt)
library(scatterplot3d)
library(copula)
```

y para realizar las ilustraciones es necesario una semilla aleatoria `set.seed(1)`.

### 5.2. Cópulas Elípticas

En cuanto a las cópulas elípticas, el paquete *copula* tiene incorporada la clase *ellipCopula*, en la cual es necesario especificar con qué familia se va a generar la cópula (*family*), o directamente con *normalCopula* y *tCopula*, que, como sus nombres lo dicen, corresponden a las cópulas normal y  $t$ , respectivamente. Entre los componentes principales de estas cópulas están la dimensión (*dim*), el parámetro (*param*) y la matriz de correlación (*dispstr*). Siempre que las cópulas sean invariantes bajo transformaciones monótonas de sus marginales, su matriz de

correlación determina su estructura de dependencia; las matrices de correlación implementadas en R son auto regresivas de orden 1 (ar1), intercambiables (ex), Toeplitz (toep) y no estructuradas (un), cuyas matrices de correlación, por ejemplo para dimensión 3, son, respectivamente:

$$\begin{pmatrix} 1 & \rho_1 & \rho_1^2 \\ \rho_1 & 1 & \rho_1 \\ \rho_1^2 & \rho_1 & 1 \end{pmatrix}, \begin{pmatrix} 1 & \rho_1 & \rho_1 \\ \rho_1 & 1 & \rho_1 \\ \rho_1 & \rho_1 & 1 \end{pmatrix}, \begin{pmatrix} 1 & \rho_1 & \rho_2 \\ \rho_1 & 1 & \rho_1 \\ \rho_2 & \rho_1 & 1 \end{pmatrix}$$

$$\text{y} \begin{pmatrix} 1 & \rho_1 & \rho_2 \\ \rho_1 & 1 & \rho_3 \\ \rho_2 & \rho_3 & 1 \end{pmatrix}$$

donde los  $\rho_i$ 's son parámetros de dispersión; así, dependiendo de cuál matriz de correlación se escoja, se debe introducir un valor o un vector en el componente principal parámetro (param).

Para ilustrar la construcción de este tipo de cópulas, a continuación se muestra un código que crea una cópula trivariada normal y una  $t$  con diferentes matrices de correlación:

```
CopulaElipNormal<-ellipCopula(family="normal",
dim=3,dispstr="ar1",param=0.7)
CopulaElipT<-tCopula(dim=3,dispstr="un",
param=c(0.3,0.5,0.8),df=2)
```

En esta última se puede observar un componente adicional (df), que corresponde a los grados de libertad de la cópula  $t$ .

### 5.3. Cópulas Arquimedianas

Para la construcción de cópulas Arquimedianas, el paquete copula tiene implementadas la clase general *arcmCopula*, en la cual se debe especificar la familia por utilizar, como las mostradas anteriormente: Clayton, Frank y Gumbel, y además hay otras, como Joe y amh. Además, otros componentes principales de esta clase son la dimensión (dim) y el valor del parámetro de ella (param), por ejemplo:

```
CopulaGumbel1<-arcmCopula(family="gumbel",
dim=3,param=2)
```

en donde se está generando una cópula trivariada Arquimediana de la familia Gumbel con parámetro  $\alpha = 2$ . Es necesario tener en cuenta que para dimensiones mayores a 2 solo se admiten asociaciones positivas.

De manera alternativa, se pueden crear cópulas de esta clase llamando directamente las familias antes nombradas de las cópulas Arquimedianas con *claytonCopula()*, *frankCopula()*, *gumbelCopula()*, *joeCopula()* y *amhCopula*. Así, para la cópula anterior se usaría el código:

```
CopulaGumbel2<-gumbelCopula(2,dim=3)
```

### 5.4. Funciones de densidad y de distribución

En ocasiones puede ser de utilidad o interés acceder a la expresión simbólica de la función de densidad de probabilidad (df) o de la función de distribución acumulada (c.d.f), lo cual en R es muy simple; las funciones de distribución acumulada para los ejemplos anteriores (CopulaGumbel1 y CopulaGumbel2), que son las mismas, dado que son diferentes formas de generar la misma cópula, están dadas por:

```
CopulaGumbel1@exprdist$cdf
```

lo cual genera:

```
exp(-((-log(u1))^alpha + (-log(u2))^alpha
+ (-log(u3))^alpha)^(1-alpha))
```

```
CopulaGumbel2@exprdist$cdf
```

que también genera:

```
exp(-((-log(u1))^alpha + (-log(u2))^alpha
+ (-log(u3))^alpha)^(1-alpha))
```

Ahora, las funciones de densidad de probabilidad para ambas cópulas se generan con los códigos:

```
CopulaGumbel1@exprdist$pdf
```

```
CopulaGumbel2@exprdist$pdf
```

los cuales generan

```
(exp(-((-log(u1))^alpha + (-log(u2))^alpha +
(-log(u3))^alpha)^(1/alpha)) *
(((log(u1))^alpha + (-log(u2))^alpha +
(-log(u3))^alpha)^(1/alpha) - 1)
* ((1/alpha) * ((-log(u3))^(alpha - 1)
* (alpha * (1/u3)))))) *
(((log(u1))^alpha + (-log(u2))^alpha
+ (-log(u3))^alpha)^(1/alpha) - 1)
* ((1/alpha) * ((-log(u2))^(alpha - 1)
* (alpha * (1/u2)))))) - exp(-((-log(u1))^alpha
+ (-log(u2))^alpha + (-log(u3))^alpha)^(1/alpha))
* (((log(u1))^alpha + (-log(u2))^alpha +
(-log(u3))^alpha)^(1/alpha) - 1) - 1)
* (((1/alpha) - 1) * ((-log(u3))^(alpha - 1)
* (alpha * (1/u3)))) * ((1/alpha) *
((-log(u2))^(alpha - 1) * (alpha * (1/u2))))))
```

```
* (((-log(u1))^alpha + (-log(u2))^alpha +
(-log(u3))^alpha)^(1/alpha) - 1) * ((1/alpha)
* ((-log(u1))^(alpha - 1) * (alpha * (1/u1))))))
- exp(-((-log(u1))^alpha + (-log(u2))^alpha
+ (-log(u3))^alpha)^(1/alpha))
* (((-log(u1))^alpha + (-log(u2))^alpha
+ (-log(u3))^alpha)^(1/alpha) - 1)
* ((1/alpha) * ((-log(u2))^(alpha - 1)
* (alpha * (1/u2)))))) * (((-log(u1))^alpha
+ (-log(u2))^alpha
+ (-log(u3))^alpha)^(1/alpha) - 1)
* (((1/alpha) - 1) * ((-log(u3))^(alpha - 1)
* (alpha * (1/u3)))))) * ((1/alpha)
* ((-log(u1))^(alpha - 1) * (alpha * (1/u1))))))
- (exp(-((-log(u1))^alpha + (-log(u2))^alpha
+ (-log(u3))^alpha)^(1/alpha))
* (((-log(u1))^alpha + (-log(u2))^alpha
+ (-log(u3))^alpha)^(1/alpha) -
1) * ((1/alpha) * ((-log(u3))^(alpha - 1)
* (alpha * (1/u3)))))) * (((-log(u1))^alpha
+ (-log(u2))^alpha
+ (-log(u3))^alpha)^(1/alpha) - 1)
* (((1/alpha) - 1) * ((-log(u2))^(alpha - 1)
* (alpha * (1/u2)))))) * ((1/alpha) *
((-log(u1))^(alpha - 1) * (alpha * (1/u1))))))
- exp(-((-log(u1))^alpha + (-log(u2))^alpha
+ (-log(u3))^alpha)^(1/alpha)) *
(((1/alpha)^(alpha - 1) * (-log(u2))^(alpha - 1)
+ (-log(u3))^alpha)^(1/alpha)) *
(((1/alpha)^(alpha - 1) * (-log(u2))^(alpha - 1)
+ (-log(u3))^alpha)^(1/alpha) - 1)
* (((1/alpha) - 1) * ((-log(u2))^(alpha - 1)
* (alpha * (1/u2)))))) * ((1/alpha)
* ((-log(u1))^(alpha - 1) * (alpha * (1/u1))))))
- exp(-((-log(u1))^alpha + (-log(u2))^alpha
+ (-log(u3))^alpha)^(1/alpha)) *
(((1/alpha)^(alpha - 1) * (-log(u2))^(alpha - 1)
+ (-log(u3))^alpha)^(1/alpha) - 1)
* (((1/alpha) - 1) * ((-log(u2))^(alpha - 1)
* (alpha * (1/u2)))))) * ((1/alpha)
* ((-log(u1))^(alpha - 1) * (alpha * (1/u1))))))
```

Para clarificar y comprobar el resultado de la distribución de la cópula con la tabla 1 es necesario generar una cópula bivariada (dim = 2), como, por ejemplo, una cópula Frank:

```
CopFrank<-frankCopula(5,dim=2)
CopFrank@exprdist$cdf
-1/alpha * log(1 + exp(-log((exp(-alpha * u1)
- 1)/(exp(-alpha) - 1)) + -log((exp(-alpha * u2)
- 1)/(exp(-alpha) - 1)))) * (exp(-alpha) - 1))
```

donde se observa que R no arroja las funciones simplificadas, sino en su forma general para la dimensión requerida, que al reducirla es equivalente a la presentada en la tabla 1:

$$C_{\alpha}(u, v) = \frac{-1}{\alpha} \ln \left[ 1 + \frac{(e^{-\alpha u} - 1)(e^{-\alpha v} - 1)}{e^{-\alpha} - 1} \right]$$

De esta forma es fácil deducir, por ejemplo, que para una dimensión de tamaño  $n$  la función de distribución de una cópula Frank está dada por:

$$C_{\alpha}(u_1, u_2, \dots, u_n) = \frac{-1}{\alpha} \ln \left[ 1 + \frac{\prod_{i=1}^n (e^{-\alpha u_i} - 1)}{e^{-\alpha} - 1} \right]$$

### 5.5. Clase mvdc

La clase *mvdc* permite construir distribuciones multivariadas con marginales dadas utilizando cópulas, para lo cual es necesario predeterminar una cópula, como en los ejemplos anteriores. Además, se especifican las funciones de distribución marginales, que pueden estar implementadas en R, o se las puede haber determinado previamente (margins), y estableciendo además cuáles son los parámetros de estas marginales (paramMargins). Por ejemplo, el siguiente código permite construir un objeto *mvdc* que representa una distribución trivariada con una cópula Gumbel y marginales normales estándar:

```
MvdcGumbelNorm1<-mvdc(copula=CopulaGumbel1,
margins=c("norm","norm","norm"),
paramMargins=list(list(mean=0,sd=1),
list(mean=0,sd=1),list(mean=0,sd=1)))
```

De igual forma, puede ser necesario generar números aleatorios y su evaluación en la función de densidad o de distribución de una cópula, lo cual se obtiene con el siguiente código:

```
rCopulaGumbel2<-rcopula(CopulaGumbel2,5)
rCopulaGumbel2
[ ,1] [ ,2] [ ,3]
[1,] 0.02124414 0.2654165 0.4105957
[2,] 0.48518769 0.5380269 0.3326853
[3,] 0.65966524 0.2361050 0.5309277
[4,] 0.22481031 0.5348209 0.2579182
[5,] 0.51056608 0.6068635 0.7049956
cbind(dcopula(CopulaGumbel2,rCopulaGumbel2),
pcopula(CopulaGumbel2,rCopulaGumbel2))
[ ,1] [ ,2]
[1,] 0.7922743 0.01545534
[2,] 1.9243196 0.23328332
[3,] 0.8251597 0.19589003
[4,] 1.6269346 0.12113929
[5,] 2.3059454 0.40353391
```

Para realizar esto con objetos *mvdc* solamente se aplica la función cuantíl a números aleatorios de la cópula en cada marginal, así:

```
randommvdc<-rmvdc(MvdcGumbelNorm1,5)
randommvdc
[ ,1] [ ,2] [ ,3]
[1,] -0.642030846 -1.2214252 -1.78578885
[2,] 0.007702543 0.3108595 0.73128241
[3,] 0.460719585 -0.7623861 -0.63856114
[4,] -0.963609021 -0.6735156 0.02300873
[5,] 1.254813002 1.5327896 1.16942591
cbind(dmvdc(MvdcGumbelNorm1,randommvdc),
pmvdc(MvdcGumbelNorm1,randommvdc))
[ ,1] [ ,2]
```

```
[1,] 0.01780205 0.01524604
[2,] 0.08583419 0.41647241
[3,] 0.03279106 0.12871690
[4,] 0.05032958 0.09453390
[5,] 0.08451204 0.83364501
```

### 5.6. Gráficas

En el campo de la estadística, las gráficas son una potente herramienta de análisis; por ejemplo, para graficar datos de cópulas trivariadas se puede utilizar el paquete `scatterplot3d`. El siguiente código re-

presenta 200 puntos aleatorios de la anterior cópula

```
Gumbel trivariada con marginales normales estándar:
CopulaGumbel1<-archmCopula(family="gumbel",
dim=3,param=2)
MvdcGumbelNorm1<-mvdc(copula=CopulaGumbel1,
margins=c("norm","norm","norm"),
paramMargins=list(list(mean=0,sd=1),
list(mean=0,sd=1),list(mean=0,sd=1)))
randommvdc<-rmvdc(MvdcGumbelNorm1,200)
scatterplot3d(randommvdc)
```

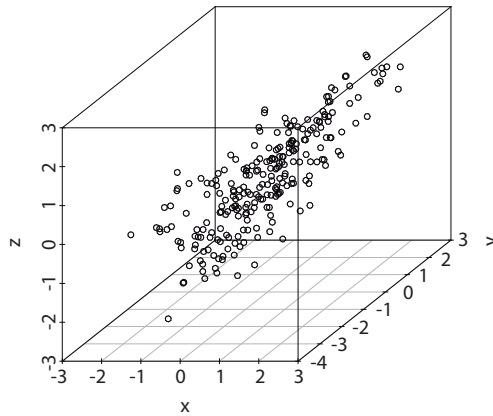


Figura 1. Puntos aleatorios de una Cópula Gumbel trivariada con marginales Normales estándar.

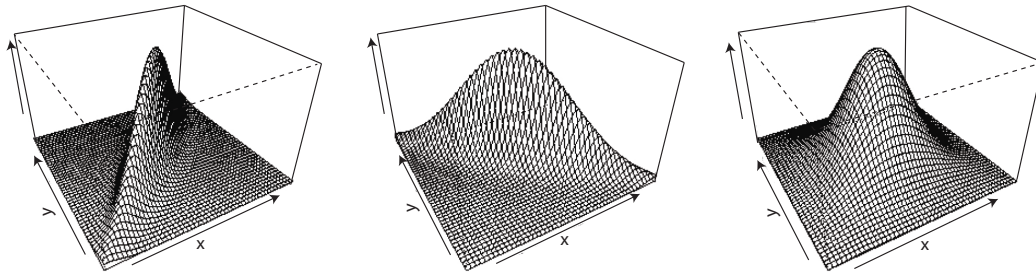


Figura 2. Perspectivas de las funciones de densidad de una cópula *t* biviada con marginales normales estándar, con  $\alpha = 0,9, -0,9, 0,1$ .

Para los objetos `cópula` y `mvdc`, el paquete `copula` tiene implementados métodos para dibujar representaciones de perspectiva y contorno para las funciones de distribución y densidad; estas funciones son `persp` y `contour`. El siguiente código permite representar las funciones de distribución y densidad de una cópula biviada con marginales normales estándar tanto en contorno como en perspectiva para un `param` de 0.9:

```
CopulaT<-ellipCopula(family="t",dim=2,
dispstr="ex",param=0.9)
```

```
MvdcCopulaT<-mvdc(copula=CopulaT,
margins=c("norm","norm"),
paramMargins= list(list(mean=0,sd=1),
list(mean=0,sd=1)))
contour(MvdcCopulaT, dmvdc,
xlim = c(-2.5, 2.5), ylim = c(-2.5, 2.5))
persp(MvdcCopulaT, dmvdc, xlim = c(-2.5, 2.5),
ylim = c(-2.5, 2.5))
contour(MvdcCopulaT, pmvdc, xlim = c(-2, 7.5),
ylim = c(-2, 7.5))
persp(MvdcCopulaT, pmvdc, xlim = c(-5, 5),
ylim = c(-5, 5))
```

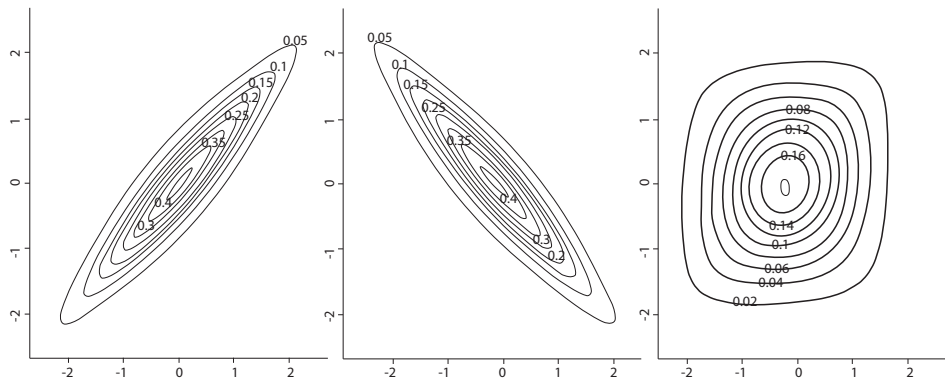


Figura 3. Contornos de las funciones de densidad de una cópula  $t$  biviada con marginales normales estándar, con  $\alpha = 0,9, -0,9, 0,1$ .

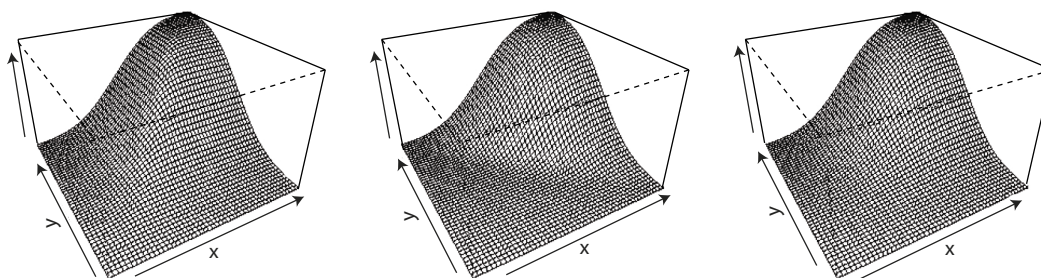


Figura 4. Perspectivas de las distribuciones de una cópula  $t$  biviada con marginales normales estándar, con  $\alpha = 0,9, -0,9, 0,1$ .

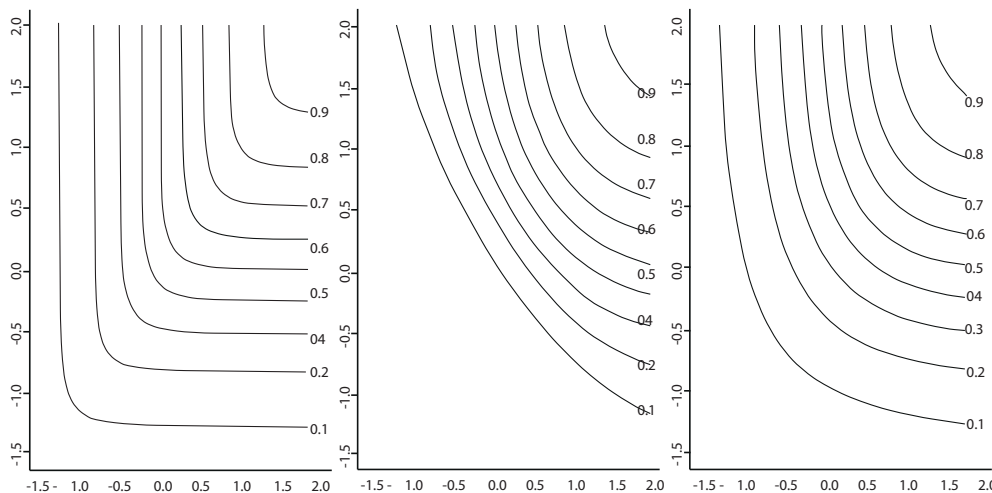


Figura 5. Contornos de las distribuciones de una cópula  $t$  biviada con marginales normales estándar, con  $\alpha = 0,9, -0,9, 0,1$ .

## 6. Comentario y conclusiones

Se espera que esta revisión del paquete copula, del software estadístico R, sea de mucha ayuda para las personas que tienen pocos conocimientos del R; el uso de este paquete se debería implementar en los cursos básicos de estadística, en especial, en el curso de estadística multivariada, ya que la estimación de distribuciones conjuntas multivariadas es un campo inexplorado o con soluciones muy delimitadas, y en estos casos el uso de cópulas puede ser de vital importancia en la estimación.

La construcción de distribuciones multivariadas es muy versátil, ya que no requiere que las marginales univariadas, sean del mismo tipo, lo que permite que al momento de utilizarlo en problemas prácticos sea muy aplicable.

La estructura de dependencia utilizada en cópulas de la familia elíptica es más amplia que la de la familia Arquimediana en dimensiones superiores a 2, ya que al momento de analizar o establecer la correlación o dependencia entre cada par de variables, las cópulas Arquimedianas solo permiten establecer,



ya sea a partir del  $\tau$  de Kendall o del  $\alpha$ , la misma dependencia para cada par de variables, contrario a lo que pasa con las cópulas elípticas, que permiten diferentes matrices de correlación, algunas de ellas con al menos dos correlaciones diferentes.

Por último, cabe destacar que el acceso gratuito al software estadístico R ha facilitado el uso de las cópulas y, por lo tanto, muchos campos se han visto beneficiados; aunque las bases teóricas de las cópulas no son muy fáciles de utilizar o no son muy amigables, como lo afirmaba [17], ahora se encuentran al alcance de todos.

## Referencias

- [1] Escarela, H. M. Mena & A. Castillo-Morales, "A Flexible Class of Parametric Transition Regression Models Based on Copulas: Application to Poliomyelitis Incidence", *Stat Methods Med Res*, vol. 15, no. 6, pp. 593-609, December 2006.
- [2] E. W. Frees & E. A. Valdez, "Understanding Relationships Using Copulas". *North American Actuarial Journal*, vol. 2, no. 1, pp. 1-25, 1998.
- [3] E. W. Frees & P. Wang, "Credibility Using Copulas". *North American Actuarial Journal*, vol. 9, no. 2, pp. 31-48, 2005.
- [4] T. Ané, "Dependence structure and risk measure", *The Journal of Business*, vol. 76, no. 3, pp. 411-438, 2003.
- [5] U. Cherubini, E. Luciano & W. Vecchiato, *Copula Methods in Finance*. John Wiley and Sons, 2004.
- [6] G. Escarela & J. F. Carriere, "Fitting Competing Risks with an Assumed Copula", *Stat Methods Med Res*, vol. 12, no. 4, pp. 333-349, August 2003.
- [7] C. Genest, & A. C. Favre, "Everything you always wanted to know about Copula Modeling but were afraid to ask", *Journal of Hydrologic Engineering*, vol. 12, no. 4, pp. 347-368, 2007.
- [8] R. Nelsen, *An Introduction to Copulas*, 2 edition. New York: Springer; 2006.
- [9] C. Genest & L-P. Rivest, "Statistical Inference Procedures for Bivariate Archimedean Copulas", *Journal of the American Statistical Association*, vol. 88, no. 423, pp. 1034-1043, sep. 1993.
- [10] C. Genest & J. MacKay, "The joy of copulas: bivariate distributions with uniform marginals". *The American Statistician*, vol. 40, no. 4, pp. 280-283, 1986.
- [11] J. Yan, "Enjoy the Joy of Copulas: With a Package Copula", *Journal of Statistical Software*, vol. 21, no. 4, pp. 1-21; Oct 2007.
- [12] P. Trivedi & D. Zimmer, *Copula Modeling: An Introduction for Practitioners*. World Scientific Publishing, 2007.
- [13] P. Embrechts, F. Lindskog & A. McNeil, *Modelling Dependence with Copulas and Applications to Risk Management*, Department of Mathematics ETHZ, Zurich, 2001; disponible en <http://www.math.ethz.ch/embrecht/ftp/copchapter.pdf>.
- [14] G. Escarela & A. Hernández, "Modelado de parejas aleatorias usando cópulas", *Revista Colombiana de Estadística*, vol. 32, no. 1, pp. 1-58, 2009.
- [15] C. Lopera, M. Jaramillo & N. Acosta, "¿Cuándo inicia la enfermedad de Alzheimer? Kaplan Meier versus Turnbull: Una aplicación a datos con censura arbitraria". *Revista Colombiana de Estadística*, vol. 35, no. 2, pp. 239-254, 2012.
- [16] G. Szegö, "Measures of risk". *Journal of Banking and Finance*, vol. 26, pp. 1253-1272, 2002.
- [17] W. Díaz, *Contribuciones a la dependencia y dimensionalidad de cópulas*, Tesis de Doctorado, Universidad de Barcelona, Facultad de Biología, Departamento de Estadística, 2012.