Simulación ITAM
28 de septiembre de 2021 2021-II

Tarea 3.

La fecha de entrega es el **1 de octubre de 2021**.

Lectura

- Salmon, F. (2009) Recipe for Disaster: The Formula That Killed Wall Street, Wired
- Arturo Ederly: Cópulas y variables aleatorias, una introducción

Problemas

1. Obtener una muestra aleatoria de 5,000 observaciones del vector $X = (X_1, X_2, X_3, X_4)$ donde $X_1 \sim \mathcal{N}(4, 9)$, $X_2 \sim$ **Bernoulli** (0.6), $X_3 \sim Beta(2, 3)$ y $X_4 \sim Gamma(3, 2)$. Considerar la siguiente estructura de correlación entre las variables (en términos de la correlación de Spearman):

   - $cor(X_1, X_3) = -0.7$
   - $cor(X_2, X_4) = 0.4$
   - $cor(X_1, X_4) = 0.5$
   - $cor(X_2, X_3) = 0.2$
   - $cor(X_i, X_j) = 0$, para las correlaciones no dadas.

   Con la muestra, hacer los histogramas de las funciones marginales y corroborar que tienen la distribución considerada, así como la estructura de dependencia dada.

   Solución.

   Ya que contamos con una matriz de correlación, podemos utilizar una cópula Gaussiana para generar la muestra solicitada. Como los datos ya están dados en términos de correlación, la matriz Σ está dada por

$$\Sigma = \begin{pmatrix} 1 & 0 & -0.7 & 0.5 \\ 0 & 1 & 0.2 & 0.4 \\ -0.7 & 0.2 & 1 & 0 \\ 0.5 & 0.4 & 0 & 1 \end{pmatrix}$$

   Entonces el procedimiento es:

a) Obtener un vector $Z \sim N(0, \Sigma)$

b) obtener un vector $U \sim (\Phi(Z_1), \Phi(Z_2), \Phi(Z_3), \Phi(Z_4))$

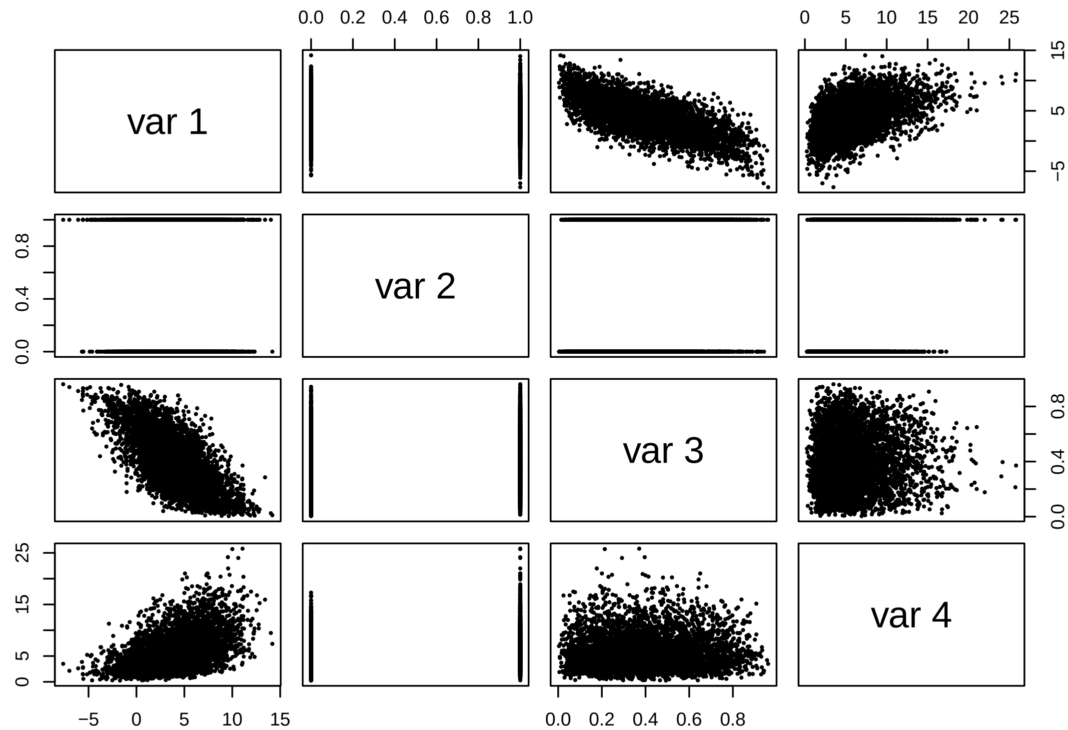
c) Obtener el vector $W \sim (F_1^{-1}(U_1), F_2^{-1}(U_2), F_3^{-1}(U_2), F_4^{-1}(U_4))$

Un ejercicio similar a este vimos en uno de los laboratorios.

```
library(copula)
cn4 <- normalCopula(c(0,-0.7,0.5,0.2,0.4,0),dim=4,dispstr = "un")
set.seed(100) #fija una semilla
U <- rCopula(5000,cn4) #Genera la muestra aleatoria
W <- cbind(qnorm(U[,1],mean=4,sd=3),
qbinom(U[,2],size=1,prob=0.6),
qbeta(U[,3],2,3),
qgamma(U[,4],shape=3,scale=2))
head(W)
```

```
         [,1] [,2]      [,3]      [,4]
[1,] 3.555827    1 0.4295056  7.964706
[2,] 5.618034    1 0.2758444  8.101240
[3,] 1.838498    0 0.4750418  4.665784
[4,] 3.274686    1 0.4470294  5.595060
[5,] 6.085712    1 0.2766029 15.324181
[6,] 3.190979    1 0.5095977  8.232991
```

```
pairs(W,pch=16,cex=0.5)
```



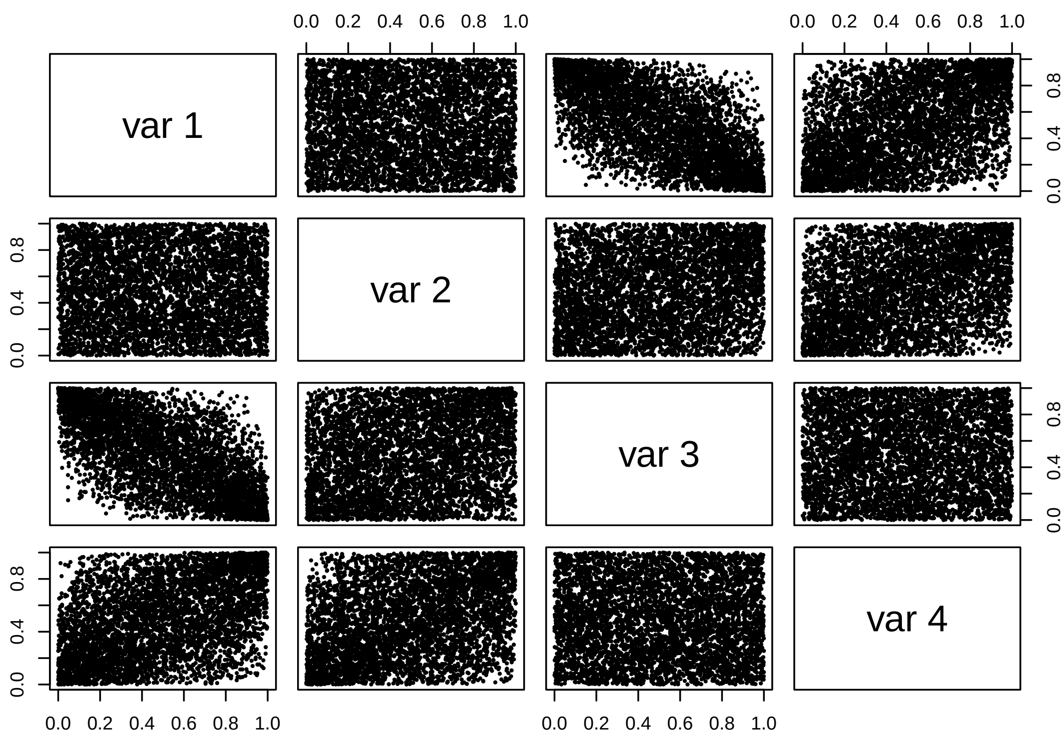

```
round(cor(W),2)
```

```
      [,1]  [,2]  [,3] [,4]
```

```
[1,]  1.00 -0.01 -0.71 0.48
[2,] -0.01  1.00  0.18 0.29
[3,] -0.71  0.18  1.00 0.00
[4,]  0.48  0.29  0.00 1.00
```

Los datos simulados de la cópula son los siguientes:

```r
pairs(U, pch=16, cex=0.5)
```

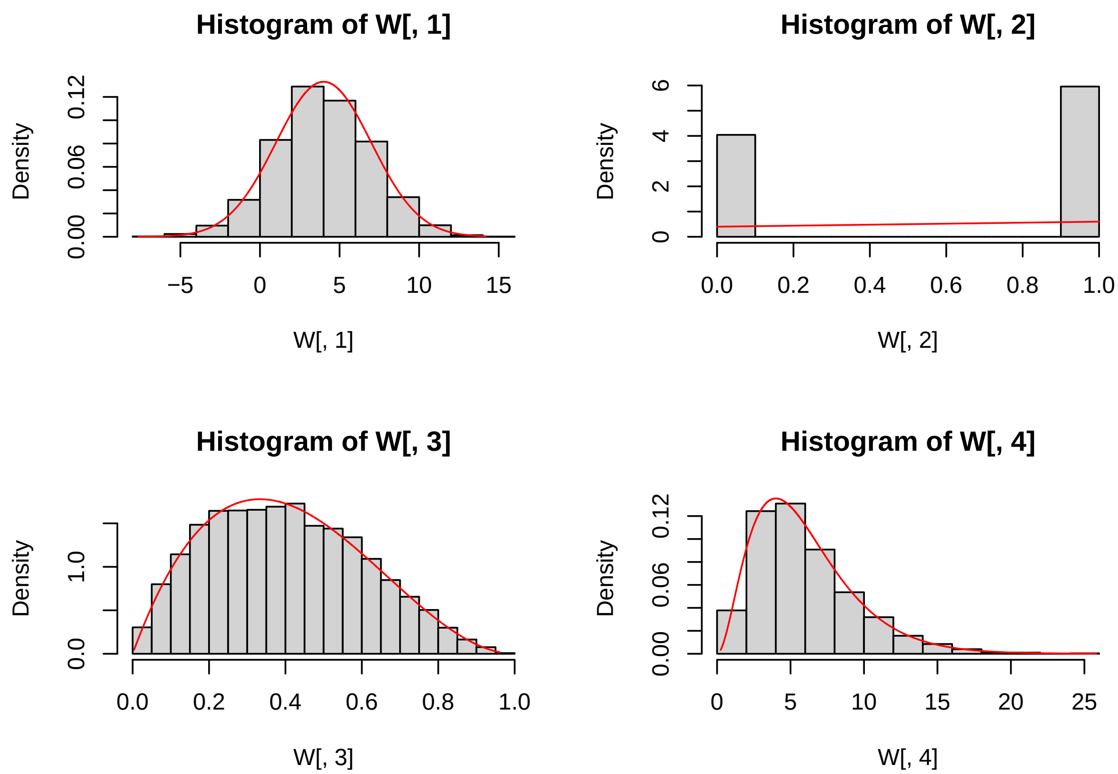



```r
round(cor(U,method = "kendall"),2)
```

```
      [,1]  [,2]  [,3] [,4]
[1,]  1.00 -0.01 -0.51 0.33
[2,] -0.01  1.00  0.13 0.25
[3,] -0.51  0.13  1.00 0.00
[4,]  0.33  0.25  0.00 1.00
```

Ahora procedemos al paso 4, para generar nuestro vector y hacemos los histogramas para ver si tienen el comportamiento deseado, y se muestra un ejemplo de los valores generados:

```r
#Grafica los histogramas y agrega densidades con las distribuciones deseadas para ver
#la aproximación
par(mfrow=c(2,2))
hist(W[,1],probability = T); points(sort(W[,1]),dnorm(sort(W[,1]),4,3),type="l",col="red")
hist(W[,2],probability = T); points(sort(W[,2]),dbinom(sort(W[,2]),1,0.6),type="l",col="red")
hist(W[,3],probability = T); points(sort(W[,3]),dbeta(sort(W[,3]),2,3),type="l",col="red")
hist(W[,4],probability = T); points(sort(W[,4]),dgamma(sort(W[,4]),shape=3,scale=2),type="l",col="red")
```




```
cor(W,method="kendall")
```

```
            [,1]        [,2]         [,3]         [,4]
[1,]  1.00000000 -0.01048706 -0.505808202  0.331958872
[2,] -0.01048706  1.00000000  0.142987645  0.251824288
[3,] -0.50580820  0.14298764  1.000000000 -0.001945189
[4,]  0.33195887  0.25182429 -0.001945189  1.000000000
```

□

2. La τ de Kendall entre X y Y es 0.55. Tanto X como Y son positivas. £Cuál es la τ entre X y $1/Y$? £Cuál es la τ de $1/X$ y $1/Y$?

   Solución.

   Hay que recordar que la τ de Kendall es una estadística basada en los rangos de la variable aleatoria (expresados en términos de las concordancias y discordancias de las observaciones) y que es invariante ante transformaciones monótonas. Entonces, partiendo de que las variables son positivas, y como $1/Y$ es monótona decreciente, el valor de la τ se mantiene, pero cambia el signo. En el caso en el que se cambian ambas variables, el valor de la estadística queda el mismo. Lo podemos comprobar con una pequeña simulación

```
X <- runif(100)
Y <- X + runif(100)
cor(X,Y,method="kendal")


[1] 0.52


cor(X,1/Y,method="kendall")


[1] -0.52


cor(1/X,1/Y,method="kendall")


[1] 0.52
```

□

3. Mostrar que cuando $\theta \to \infty$, $C^{Fr}(u_1,u_2) \to min\{u_1,u_2\}$, donde C^{Fr} es la cópula de Frank.

   Solución.

   Cuando $\theta \to \infty$, $e^{-\theta} - 1 \approx -1$, por lo que

$$
\begin{aligned}
C^{Fr}(u,v) &\approx -\frac{1}{\theta}\log[1-(e^{-\theta u}-1)((e^{-\theta v}-1)] \\
&= -\frac{1}{\theta}\log[1-(e^{-\theta(u+v)}-e^{-\theta u}-e^{-\theta v}+1)] \\
&= -\frac{1}{\theta}\log[-e^{-\theta(u+v)}+e^{-\theta u}+e^{-\theta v}] \\
&= -\frac{1}{\theta}\log[e^{-\theta u}(-e^{-\theta v}+1+e^{-\theta(v-u)})] \quad \text{si } u<v \\
&= u-\frac{1}{\theta}\log[1-e^{-\theta v}+e^{-\theta(v-u)}] \\
&\to u \quad \text{cuando } \theta\to\infty
\end{aligned}
$$

   Y el límite es simétrico, por lo que $\theta \to \infty$, $C^{Fr}(u,v) \to min\{u,v\}$.

□

4. Consideren la cópula de Clayton. Mostrar que converge a la cópula de comonotonicidad cuando $\theta \to \infty$. [Hint: usen la regla de l'Hópital considerando que la cópula de Clayton se puede escribir como $\exp\{\log(u_1^{-\theta}+u_2^{-\theta}-1)/\theta\}$ para θ positivo.]

   Solución.

Sea $u < v$ para $u, v \in (0,1)$. Entonces $log(u) > log(v)$ y por lo tanto $\theta(log(v) - \theta log(u) < 0$. Siguiendo el hint, notemos que

$$\begin{aligned} log\left(u^{-\theta} + v^{-\theta} - 1\right) &= log\left(e^{-\theta log(u)}(1 + e^{\theta(log(v) - log(u))} + e^{\theta log(u)})\right) \\ &= -\theta log(u) + log\left(1 + e^{\theta(log(v) - log(u))} + e^{\theta log(u)}\right) \end{aligned}$$

Así que

$$\begin{aligned} \exp\left(\frac{1}{\theta} log\left(u^{-\theta} + v^{-\theta} - 1\right)\right) &= \exp(-log(u)) \exp\left(\frac{1}{\theta} log\left(1 + e^{\theta(logv - logu)} + e^{\theta logu}\right)\right) \\ &= u \exp\left(\frac{-log(1 + e^{\theta k} + e^{\theta m})}{\theta}\right) \end{aligned}$$

donde $k = log(u) - log(v) < 0$ y $m = log(u) < 0$ Si aplicamos l'Hôpital a este cociente, tenemos

$$lim_{\theta \to \infty} \frac{-log(1 + e^{\theta k} + e^{\theta m})}{\theta} = lim_{\theta \to \infty} \frac{ke^{\theta k} + me^{\theta m}}{1 + e^{\theta k} + e^{\theta m}} = 0$$

Así que $u \exp\left(\frac{-log(1+e^{\theta k}+e^{\theta m})}{\theta}\right) \to ue^0 = u$. Como el resultado es simétrico en u y en v, se tiene que

$$lim_{\theta \to \infty} C(u, v)^C = min\{u, v\}$$

□

5. Supongan que tienen dos vectores de datos $(x_1, \ldots, x_n)$ y $(y_1, \ldots, y_n)$. Entonces la cópula empírica es la función $C : [0,1] \times [0,1] \to [0,1]$ definida por

$$C(u, v) = \frac{1}{n} \sum_{j=1}^{n} I\left(\frac{r_j}{n+1} \leq u, \frac{s_j}{n+1} \leq v\right)$$

donde $(r_1, \ldots, r_n)$ y $(s_1, \ldots, s_n)$ denotan los vectores de rangos de x y y respectivamente.

Escriban una función llamada `empCopula` que tome cuatro argumentos `u`, `v`, `xVec` y `yVec`. Pueden suponer que los valores `u`,`v` están en $[0,1]$ y que `xVec` y `yVec` son vectores numéricos que tienen la misma longitud (no vacios).

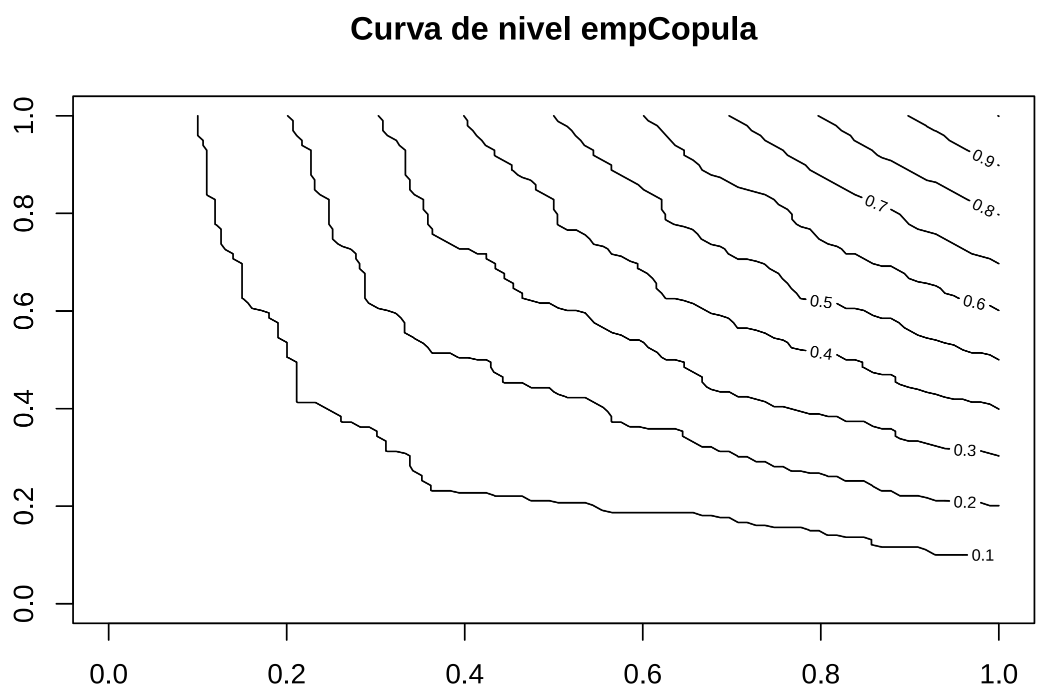
Solución.

La función que se pide es la siguiente. Es muy simple pero no está *vectorizada*, por lo que no puedo aplicar la función `outer` para poder generar un grid y hacer una gráfica.

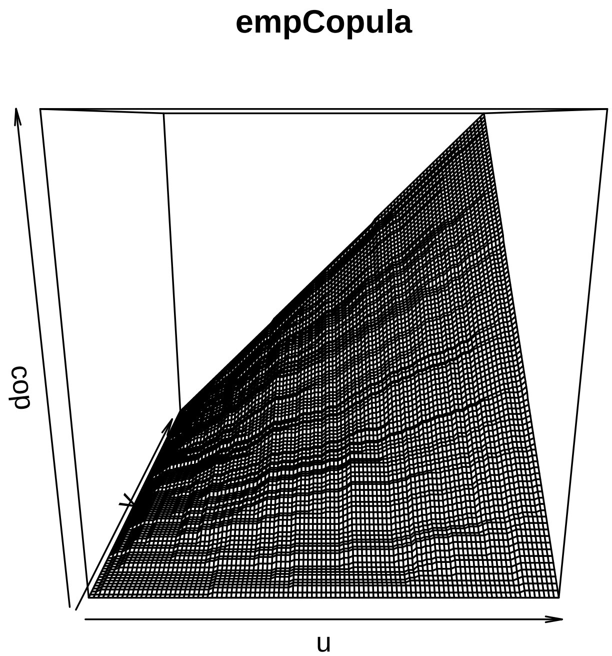
```r
empCopula <- function(u,v,xVec,yVec){
#esta función calcula la cópua empírica de un par de vectores aleatorios
#que tienen la misma longitud.
n <- length(xVec)
rx <- rank(xVec)/(n+1)
ry <- rank(yVec)/(n+1)
return(mean((rx<=u) & (ry<=v)))
}
```

Para hacer la gráfica, genero manualmente el grid para la función

```r
k <- 100
u <- v <- seq(0, 1, length = k)
xVec <- runif(200)
yVec <- rnorm(200)
cop <- matrix(numeric(),nrow=k,ncol=k)

#genera un grid para graficar:
for(uu in u)
for(vv in v)
cop[which(uu==u),which(vv==v)] <- empCopula(uu,vv,xVec,yVec)


contour(u, v, cop, main = "Curva de nivel empCopula" )
```




```r
persp(u,v,cop, main= "empCopula")
```



$\square$

6. la cópula Farlie-Gumbel-Morgenstern es $C(u,v) = uv[1+\alpha(1-u)(1-v)]$ para $|\alpha| \leq 1$. Mostrar que la densidad conjunta correspondiente $\partial^2 C(u,v)/\partial u \partial v$ es no negativa. Mostrar que C tiene marginales uniformes en $(0,1)$. Encontrar el coeficiente de correlación de Spearman y la tau de Kendall.

Solución.

Este ejercicio tiene tres partes:

- Para obtener la densidad conjunta, derivamos con respecto a cada una de las variables:

$$\begin{aligned}\frac{\partial C}{\partial u} &= v + \alpha v(1-v)[1-2u] \\ \frac{\partial^2 C}{\partial u \partial v} &= 1 + \alpha[1-2u][1-2v]\end{aligned}$$

  Como $1-2u \leq 0$ y $1-2v \leq 0$, para $0 \leq u, v \leq 1$ entonces el producto toma el valor mínimo en $u=1$ y $v=0$ o $u=0$ y $v=1$. En ese caso, para que el producto sea no negativo basta que $\alpha \leq 1$, lo cual siempre se cumple.

- Para ver que C tiene marginales uniformes, basta con evaluar C en los siguientes valores $C(1,v)$ y $C(u,1)$. Como la función es simétrica en los dos valores, basta hacer $C(1,v) = v + \alpha v(0)(1-v) = v$. Por lo tanto, las marginales son uniformes.
- Para esta parte, tenemos que resolver las ecuaciones:

$$\tau = 4\int_0^1 \int_0^1 C(u,v)dC(u,v) - 1$$

para la τ de Kendall y

$$\rho_S = 12\int_0^1 \int_0^1 C(u,v)dudv - 3$$

para la ρ de Spearman.
Haciendo primero la fórmula de la τ de Kendall, tenemos que:

$$\tau = 4\int_0^1 \int_0^1 (uv + \alpha uv(1-u)(1-v)(1+\alpha(1-2u)(1-2v))dudv - 1$$

Haciendo el producto, cada una de las integrales dobles es fácil de resolver todas son polinomiales y simétricas. Por ejemplo:

$$\int_0^1 \int_0^1 uvdudv = 1/4$$

$$\int_0^1 \int_0^1 \alpha uv(1-2u)(1-2v)dudv = \alpha/36$$

$$\int_0^1 \int_0^1 \alpha uv(1-u)(1-v)dudv = \alpha/36$$

$$\int_0^1 \int_0^1 \alpha^2 uv(1-u)(1-v)(1-2u)(1-2v)dudv = 0$$

Entonces $\tau = 4(1/4 + \alpha/18) - 1 = 1 + 2\alpha/9 - 1 = 2\alpha/9$. Para el segundo caso,

$$\rho_S = 12\int_0^1 \int_0^1 uv(1 + \alpha(1-u)(1-v))dudv - 3$$

Expandiendo los términos, nos quedan de nuevo integrales que ya hicimos, por lo que

$$\rho_S = 12(1/4 + \alpha/36) - 3 = 3 + 12\alpha/36 - 3 = \alpha/3$$

$\square$

7. Este es un ejercicio de calibración de las cópulas utilizando correlaciones de rangos. Supongan que una muestra produce un estimado de la τ de Kendall de 0.2. £Qué parámetro debe usarse para

   i. la cópula normal,
   ii. la cópula de Gumbel,
   iii. la cópula de Calyton?

   Solución.

   i. La correlación en la cópula normal – y de hecho en cualquier elíptica – debe ser establecida igual a $\rho = \sin(0.1\pi) = 0.309$.
   ii. En la cópula de Gumbel se establece $\delta = (1 - 0.2)^{-1} = 1.25$
   iii. En la cópula de Clayton establecemos $\alpha = 0.4(1 - 0.2)^{-1} = 0.5$.

   □

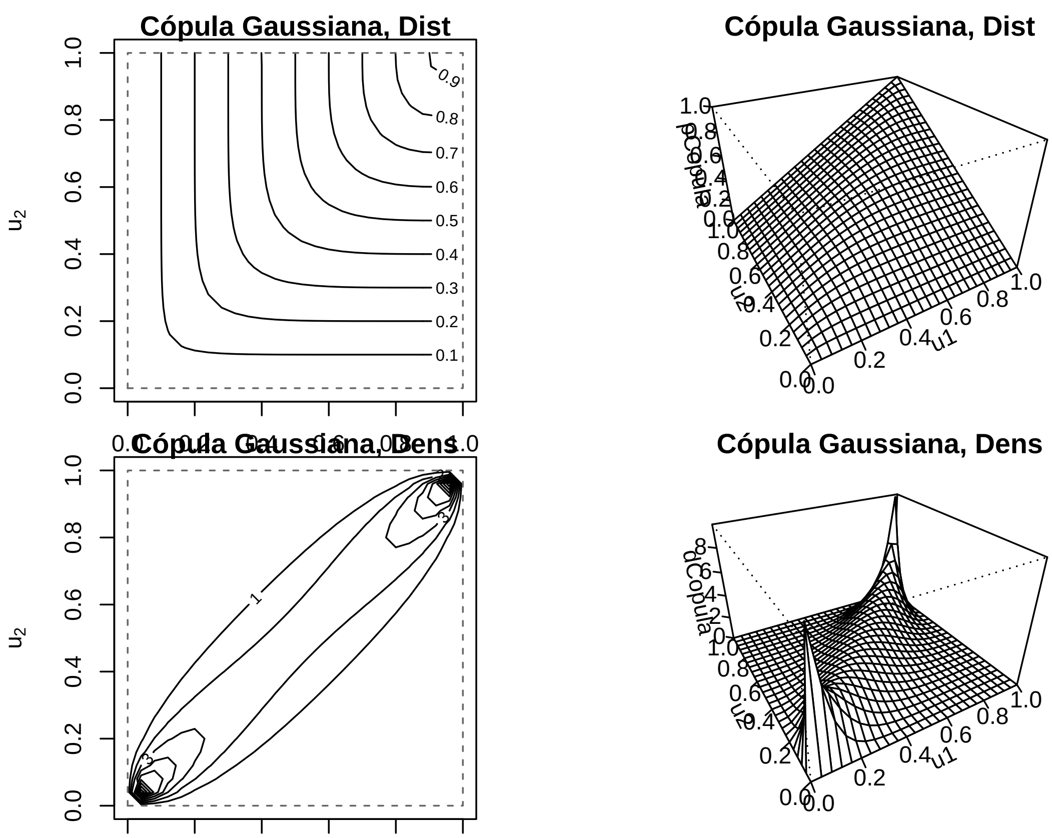
8. Usen la función `normalCopula` del paquete `copula` para crear una cópula gaussiana bidimensional con un parámetro de 0.9. Luego creen otra cópula gaussiana con parámetro de 0.2 y describan la estructura de ambas cópulas (diferencias y semejanzas).

   Solución.

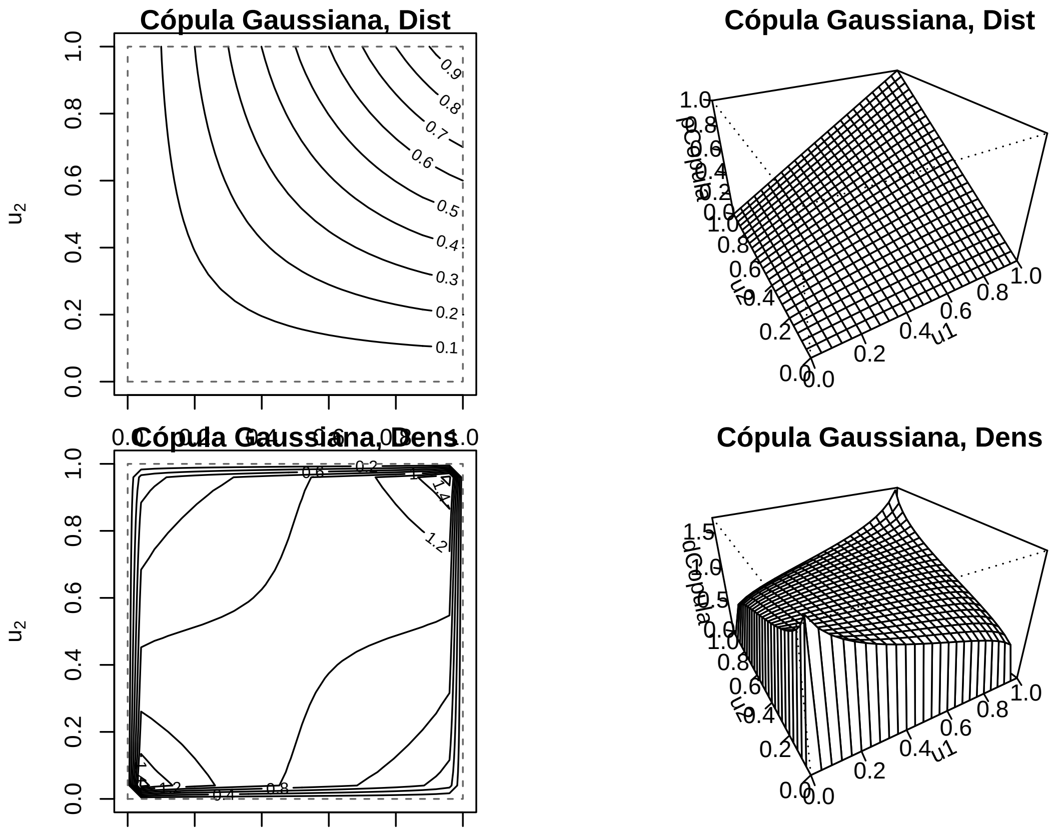
   Las siguientes gráficas ayudan a entender las diferencias entre las diferentes cópulas. Claramente la cópula con el coeficiente de correlación más alto es la que relaciona linealmente mejor a las dos variables. La dependencia lineal en la cópula gaussiana se relaciona directamente con el parámetro.

```
library(copula)
normal_0.9 <- normalCopula(param = 0.9, dim = 2)
str(normal_0.9) #despliega las características de la cópula


Formal class 'normalCopula' [package "copula"] with 8 slots
  ..@ dispstr     : chr "ex"
  ..@ getRho      :function (obj)
  ..@ parameters  : num 0.9
  ..@ param.names : chr "rho.1"
  ..@ param.lowbnd: num -1
  ..@ param.upbnd : num 1
  ..@ fullname    : chr "<deprecated slot>"
  ..@ dimension   : int 2


normal_0.2 <- normalCopula(param = 0.2, dim = 2)
str(normal_0.2 ) #despliega las características de la cópula
```

```
Formal class 'normalCopula' [package "copula"] with 8 slots
  ..@ dispstr     : chr "ex"
  ..@ getRho      :function (obj)
  ..@ parameters  : num 0.2
  ..@ param.names : chr "rho.1"
  ..@ param.lowbnd: num -1
  ..@ param.upbnd : num 1
  ..@ fullname    : chr "<deprecated slot>"
  ..@ dimension   : int 2


par(mar=c(1,2,1,1))
par(mfrow = c(2,2), pty="s")

contour(normal_0.9, pCopula, main = "Cópula Gaussiana, Dist" )
persp(normal_0.9, pCopula, main = "Cópula Gaussiana, Dist")
contour(normal_0.9, dCopula, main = "Cópula Gaussiana, Dens" )
persp(normal_0.9, dCopula, main = "Cópula Gaussiana, Dens")
```



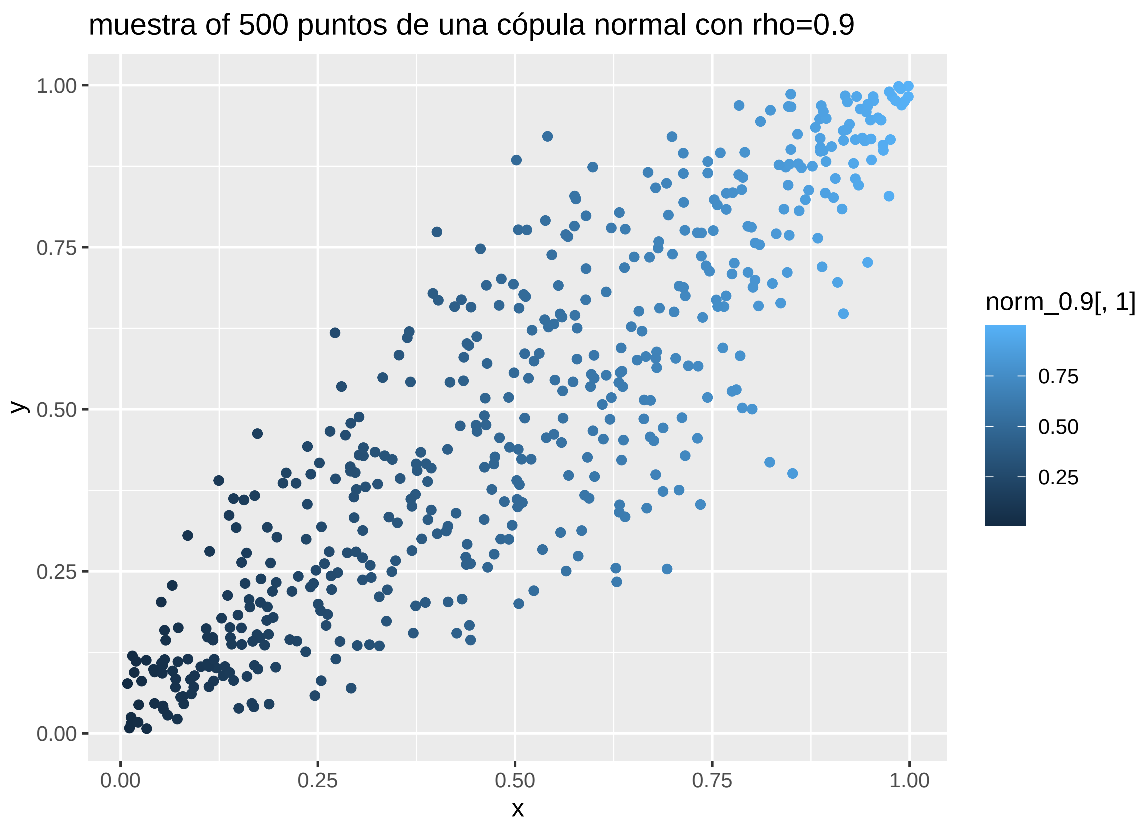

```
contour(normal_0.2, pCopula, main = "Cópula Gaussiana, Dist" )
persp(normal_0.2, pCopula, main = "Cópula Gaussiana, Dist")
contour(normal_0.2, dCopula, main = "Cópula Gaussiana, Dens" )
persp(normal_0.2, dCopula, main = "Cópula Gaussiana, Dens")
```



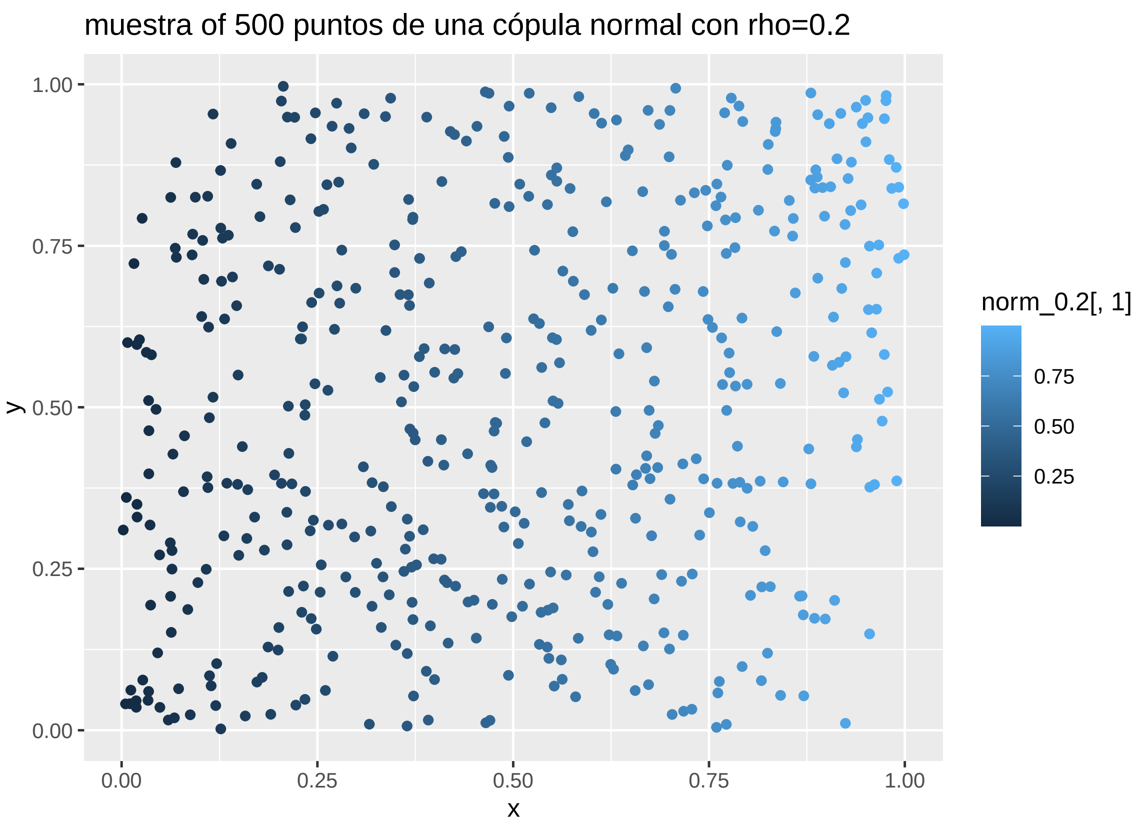
9. Usen la función `rCopula` del paquete `copula` para generar muestras de 500 puntos cuya distribución son las cópulas del ejercicio 8 anterior. Hagan una gráfica de las dos muestras. Teniendo en mente que una cópula determina la estructura de dependencia de una distribución multivariada conjunta, mirando estas gráficas, £pueden decir cuál de estas dos cópulas debe usarse para simular una distribución con una fuerte dependencia entre las marginales?

   Solución.

```
norm_0.9 <- rCopula(500, normal_0.9) #muestra de la normal con rho=0.9
norm_0.2 <- rCopula(500, normal_0.2) #muestra de la normal con rho=0.2
library(ggplot2)
plot.norm_0.9 <- qplot(norm_0.9[,1], norm_0.9[,2], colour =     norm_0.9[,1],
                                        main="muestra of 500 puntos de una cópula normal con rho=0.9", xlab = "x", ylab = "y")
plot.norm_0.9
```




```
plot.norm_0.2 <- qplot(norm_0.2[,1], norm_0.2[,2], colour = norm_0.2[,1],
                main="muestra of 500 puntos de una cópula normal con rho=0.2", xlab = "x", ylab = "y")
plot.norm_0.2
```



Claramente la cópula con parámetro 0.9 es la que relaciona densidades marginales con mayor dependencia.

□