Tarea 4

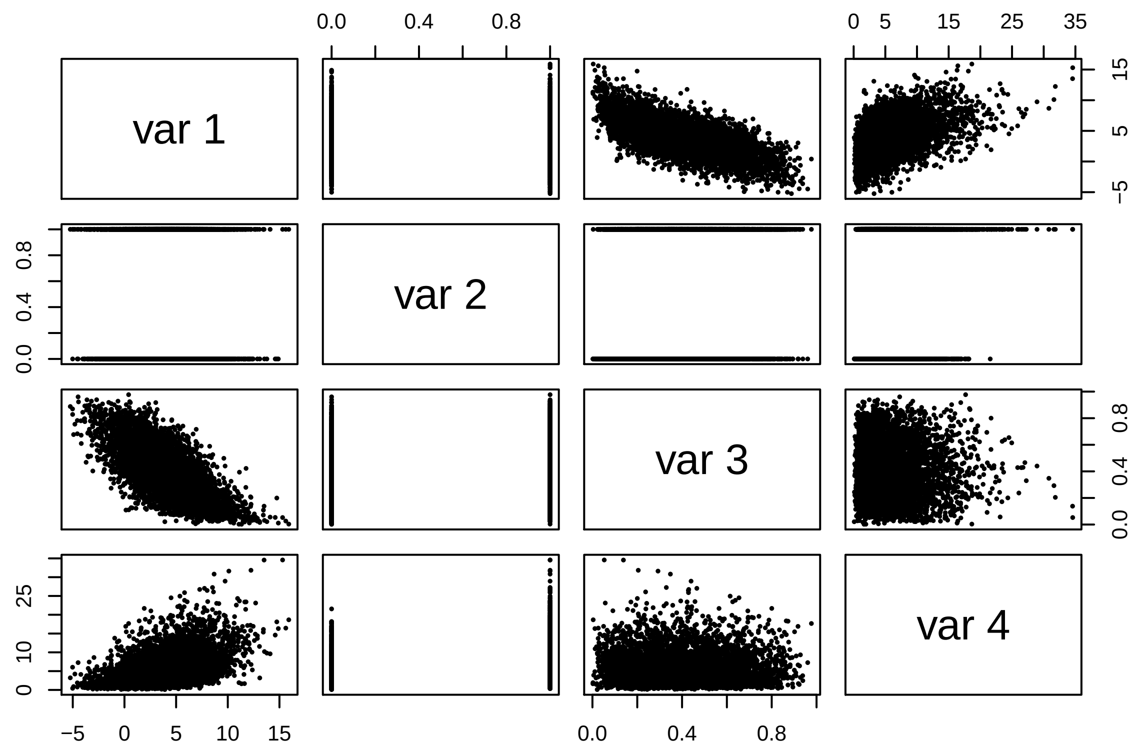
Pablo Gracia Galeana
Cesar Gonzalez Macedo
Miguel Ángel Fuentes Borboa
Roberto Antonio Yglesias Galeana

Pregunta 1

```
matriz_de_correlacion <- normalCopula(c(0,-0.7,.5,0.2,0.4,0),dim = 4,dispstr = "un")

Y <- rCopula(5000,matriz_de_correlacion)
H  <- cbind(qnorm(Y[,1],mean = 4,sd=3),qbinom(Y[,2],size = 1,prob = .6),qbeta(Y[,3],2,3),qgamma(Y[,4],s

pairs(H,pch = 16,cex = .5)
```



Pregunta 2

La τ de kendall es invariante ante transformaciones monotona. Entonces partiendo de que las variables son positivas, y como $1/Y$ es monotona decreciente, el valor de la τ se mantiene, pero cambia de signo. En el caso en el que se cambian ambas variables, el valor de la estadistica queda el mismo.

Pregunta 3

Mostrar que cuando $\theta \to \infty$, $C^{Fr}(u_1, u_2) \to min\{u_1, u_2\}$

$$C^{Fr}(u_1, u_2) = -\frac{1}{\theta}\ln(1 + \frac{(e^{-\theta u_1} - 1)(e^{-\theta u_2} - 1)}{e^{-\theta} - 1})$$

Si $u_1 < u_2$, podemos reescribir la parte dentro del logaritmo como:

$$\frac{e^{-\theta} - 1 + (e^{-\theta u_1} - 1)(e^{-\theta u_2} - 1)}{e^{-\theta} - 1} = \frac{e^{-\theta} - 1 + e^{-\theta u_1 - \theta u_2} - e^{-\theta u_1} - e^{-\theta u_2} + 1}{e^{-\theta} - 1}$$

$$= \frac{e^{-\theta} + e^{-\theta(u_1+u_2)} - e^{-\theta u_1} - e^{-\theta u_2}}{e^{-\theta} - 1} = e^{-\theta u_1}\frac{1 + e^{-\theta(u_1+u_2)} - e^{-\theta(1-u_1)} + e^{-\theta u_2}}{1 - e^{-\theta}}$$

$\therefore\ C^{Fr}(u_1, u_2) = u_1 - \frac{1}{\theta}\ln(\frac{1+e^{-\theta(u_1+u_2)}-e^{-\theta(1-u_1)}+e^{-\theta u_2}}{1-e^{-\theta}})$

$\therefore\ \lim_{\theta\to\infty} C^{Fr}(u_1, u_2) = u_1$

Análogamente, si $u_2 < u_1 \implies \lim_{\theta\to\infty} C^{Fr}(u_1, u_2) = u_2$

$\therefore$ Si $\theta \to \infty \implies \lim_{\theta\to\infty} C^{Fr}(u_1, u_2) = min\{u_1, u_2\}$

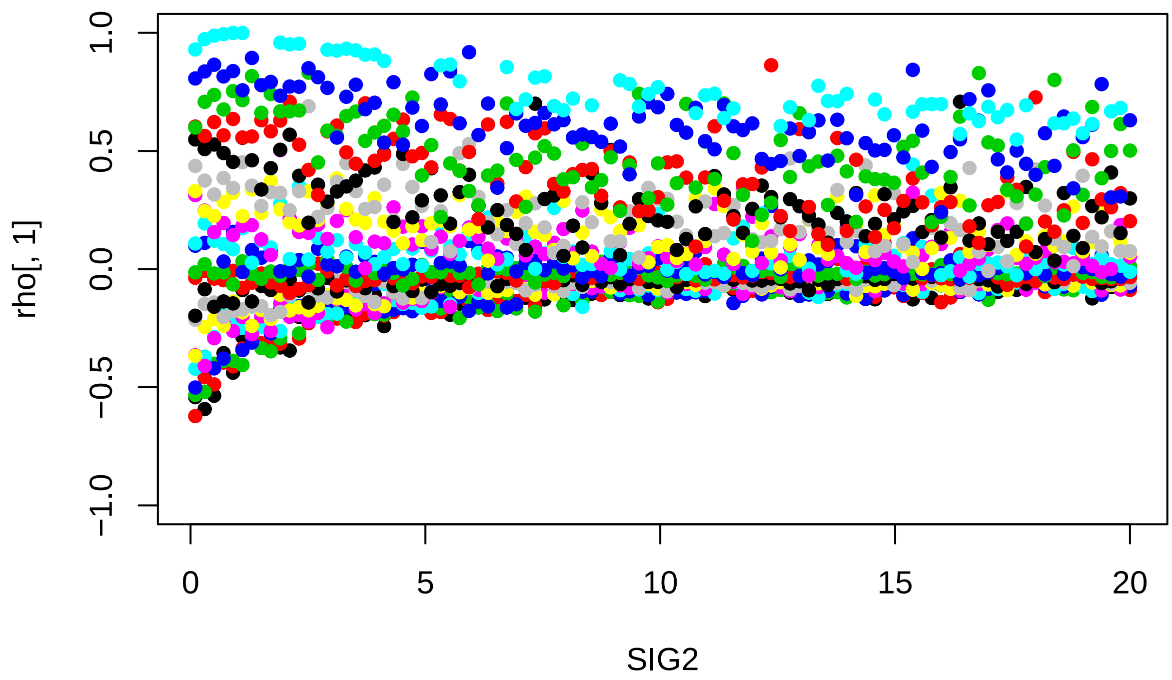
Pregunta 4

```r
SIG2 <- seq(.1,20,length.out = 100)
RHO <- seq(-1,1,length.out = 21)
rho <- matrix(numeric(),nrow=100,ncol=21)

suppressWarnings(
for(i in SIG2){
  for(j in RHO){
    Sigma <- matrix(c(1,sqrt(i)*j,sqrt(i)*j,i),nrow = 2,byrow = T)
    e <- eigen(Sigma)
    B <- e$vectors %*% diag(sqrt(e$values))%*%t(e$vectors)
    ZEXP <- exp(cbind(rnorm(500),rnorm(500))%*%B)
    rho[match(i,SIG2),match(j,RHO)] <- cor(ZEXP[,1],ZEXP[,2])
  }
}
)

plot(SIG2,rho[,1],pch = 16,col=1,ylim=c(-1,1))

for(i in 2:21)points(SIG2,rho[,i],pch=16,col=i)
```



Pregunta 5

La cópula de Clayton $C(u_1, u_2) = (u_1^{-\theta} + u_2^{-\theta} - 1)^{-\frac{1}{\theta}}$, $\theta > 0$
Si $u_2 < u_1$, entonces $u_2^{-\theta} < u_2^{-\theta} + u_1^{-\theta} - 1 < 2u_2^{-\theta}$, entonces

$$2^{-\frac{1}{\theta}} u_2 \leq (u_1^{-\theta} + u_2^{-\theta} - 1)^{-\frac{1}{\theta}} \leq u_2$$

Si $\theta \to \infty \implies 2^{-\frac{1}{\theta}} u_2 \to u_2$, entonces $u_2 \leq C(u_1, u_2) \leq u_2$
$\therefore C(u_1, u_2) \to u_2$

Si $u_2 > u_1$, se da el caso análogo donde $C(u_1, u_2) \to u_1$, si $\theta \to \infty$

$\therefore$ Si $\theta \to \infty \implies C(u_1, u_2) \to min\{u_1, u_2\}$ que es la cópula de comonotonicidad.

Pregunta 6

```
empCopula <- function(u,v,xVec,yVec){
  n <- length(xVec)
  r <- NULL
  s <- NULL
```

```
  for (i in 1:n) {
    for (j in 1:n) {
      if (xVec[j] <= xVec[i]) {
        r[i] <- r[i]+1
      }
    }
  }

  for (i in 1:n) {
    for (j in 1:n) {
      if (yVec[j] <= yVec[i]) {
        s[i] <- s[i]+1
      }
    }
  }

  for (j in 1:n) {
    if (r[j]/n+1<=u & s[j]/n+1<= v) {
      Cuv <- Cuv+1
    }
  }

  return((1/n)*Cuv)

}
```

Pregunta 7

Sea $C(u,v) = uv[1 + \alpha(1-u)(1-v)]$ para $|\alpha| < 1$

a)Mostrar que $\frac{\partial^2 C(u,v)}{\partial u \partial v}$ es no negativa.

$$\frac{\partial C(u,v)}{\partial u} = \frac{\partial}{\partial u}(uv + \alpha uv(1-u-v+uv)) = \frac{\partial}{\partial u}(uv + \alpha(uv - u^2v - v^2u - v + u^2v^2)) = v + \alpha(v - 2uv - v^2 - 2uv^2)$$

$$\frac{\partial^2 C(u,v)}{\partial u \partial v} = 1 + \alpha(1 - 2u - 2v - 4uv)$$

Ahora,
$0 \leq 2u \leq 2$, $0 \leq 2v \leq 2$, $0 \leq 4uv \leq 4$
Entonces,
$-2 \leq 4uv - 2u - 2v \leq 0$ y $-1 \leq 1 - 2u - 2v + 4uv \leq 1$
por lo que $-1 \leq \alpha(1 - 2u - 2v + 4uv) \leq 1$, de donde se deduce que $0 \leq 1 + \alpha(1 - 2u - 2v + 4uv) \leq 2$

b) Mostrar que C tiene marginales uniformes en $(0,1)$

   Por definición $C(u,v)$ cumple con la condición de que $C(u,v) = P(U \leq u, V \leq v)$

   Entonces $C(u,1) = P(U \leq u, V \leq 1) = P(U \leq u)$

   Para el caso particular donde $C(u,v) = uv[1 - \alpha(1-u)(1-v)]$, $C(u,1) = u[1 - \alpha(1-u)(1-1)] = u$, lo cuál corresponde a la distribución de una uniforme.

El caso es análogo para $C(1, v) = P(V \leq v)$

$$\therefore \text{La cópula de Farlie-Gumbel-Monstern tiene marginales uniformes en } (0, 1)$$

c)Encontrar ρ_s

$$\rho_s = 12 \int_0^1 \int_0^1 C(u,v) - uvdudv = 12 \int_0^1 \int_0^1 uv + \alpha uv(1-u)(1-v) - uvdudv$$

$$= 12 \int_0^1 \int_0^1 \alpha(uv - u^2v - uv^2 + u^2v^2)dudv = 12 \int_0^1 \alpha(\frac{v}{2} - \frac{v}{3} - \frac{v^2}{2} + \frac{v^2}{2})dv$$

$$= 12\alpha(\frac{1}{4} - \frac{2}{6} - \frac{1}{9}) = \frac{\alpha}{3}$$

d)Encontrar τ_k

$$\tau_k = 4 \int_0^1 \int_0^1 C(u,v)dC(u,v) - 1 = 4 \int_0^1 \int_0^1 uvdudv + 4 \int_0^1 \int_0^1 uv - 2u^2v - 2uv^2 + 4u^2v^2dudv$$

$$+4\alpha \int_0^1 \int_0^1 uv - u^2v - uv^2 + u^2v^2dudv + 4\alpha \int_0^1 \int_0^1 uv - 3u^2v - 3uv^2 + 9u^2v^2 + 2u^3v - 6u^3v^2 - 6u^2v^3dudv - 1$$

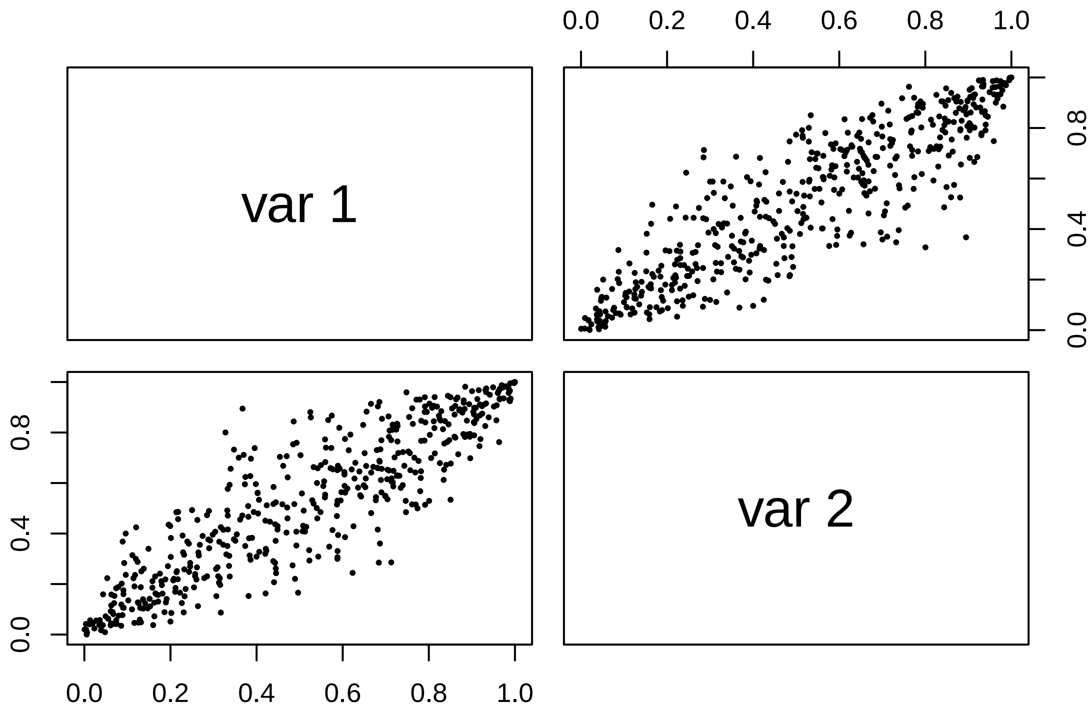
$$= 1 + \frac{\alpha}{9} + \frac{\alpha}{9} + 0 - 1 = \frac{2\alpha}{9}$$

Preguntа 8

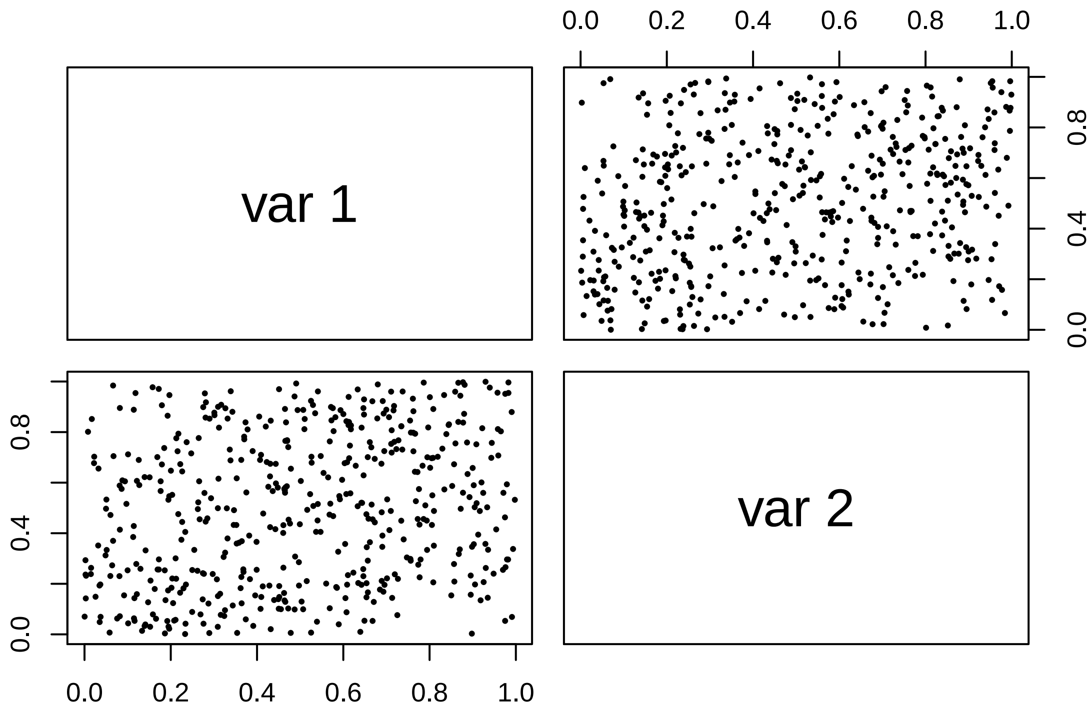
```
Cop_norm <- rCopula(500,normalCopula(sin(pi*.2/2),dispstr = "un"))
Cop_Clay <- rCopula(500,claytonCopula(2/((1-.2)/.2)))
Cop_Gumb <- rCopula(500,gumbelCopula(1/(1-.2)))
```

Pregunta 9

```
Y <- rCopula(500,normalCopula(.9,dispstr = "un"))
P <- rCopula(500,normalCopula(.2,dispstr = "un"))
pairs(Y,pch = 16,cex = .5)
```



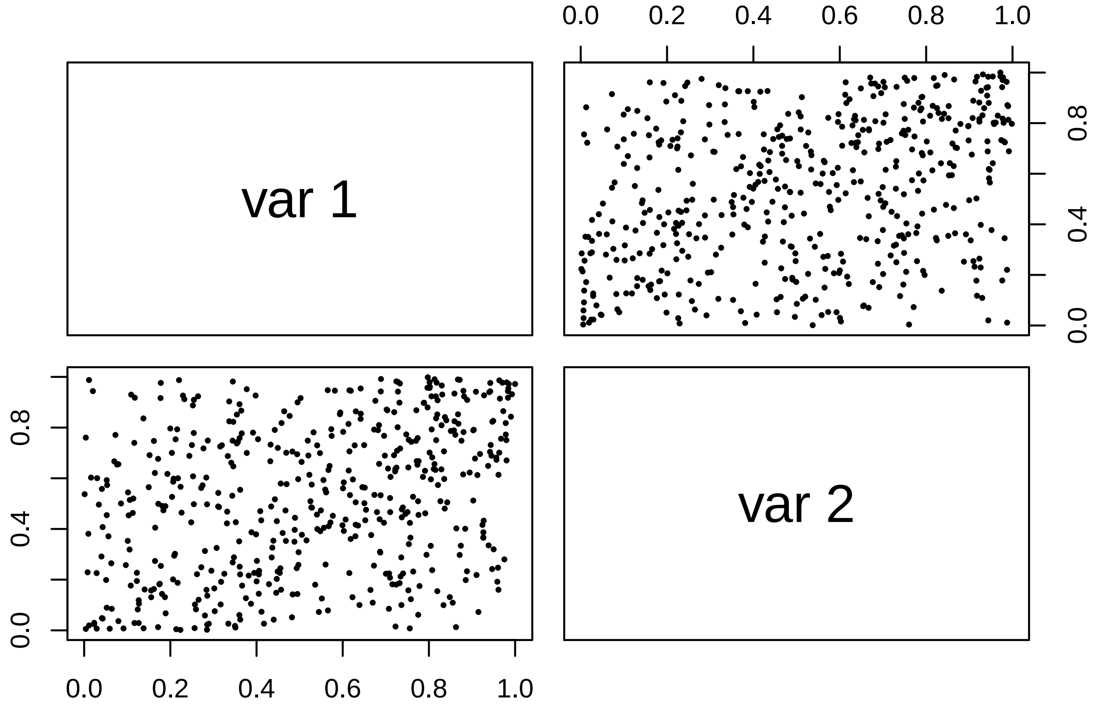

```r
pairs(P,pch = 16,cex = .5)
```



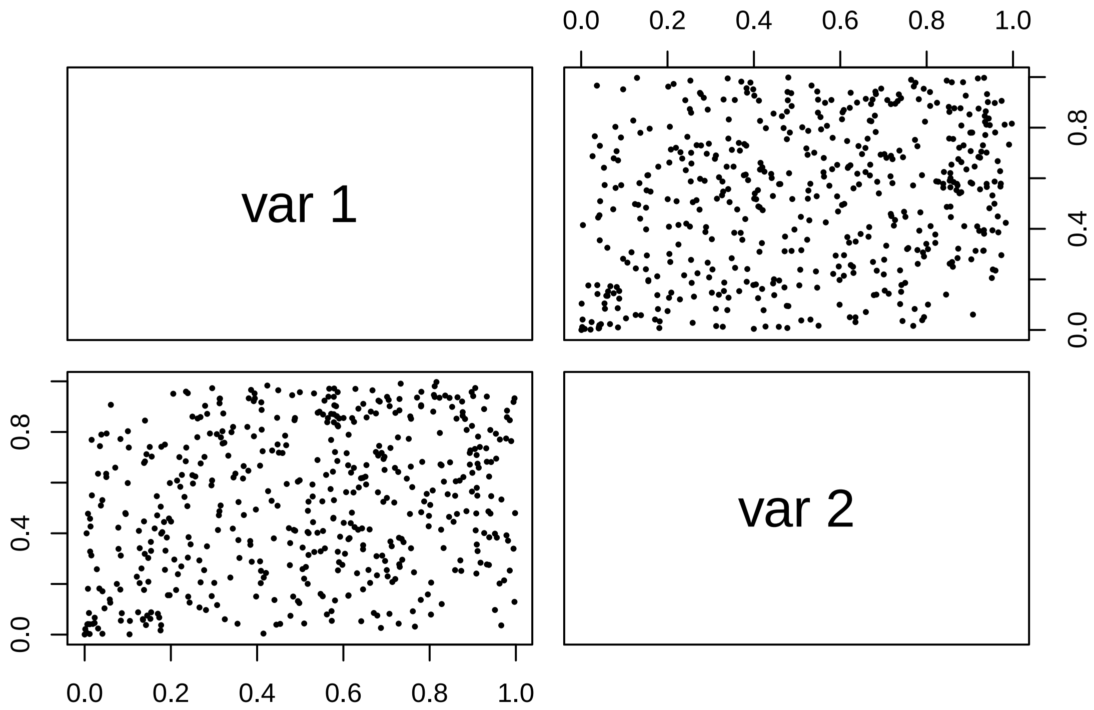
Es muy notoria la alta correlacion entre las varibles de la primera copula con respecto a la segunda.

Pregunta 10

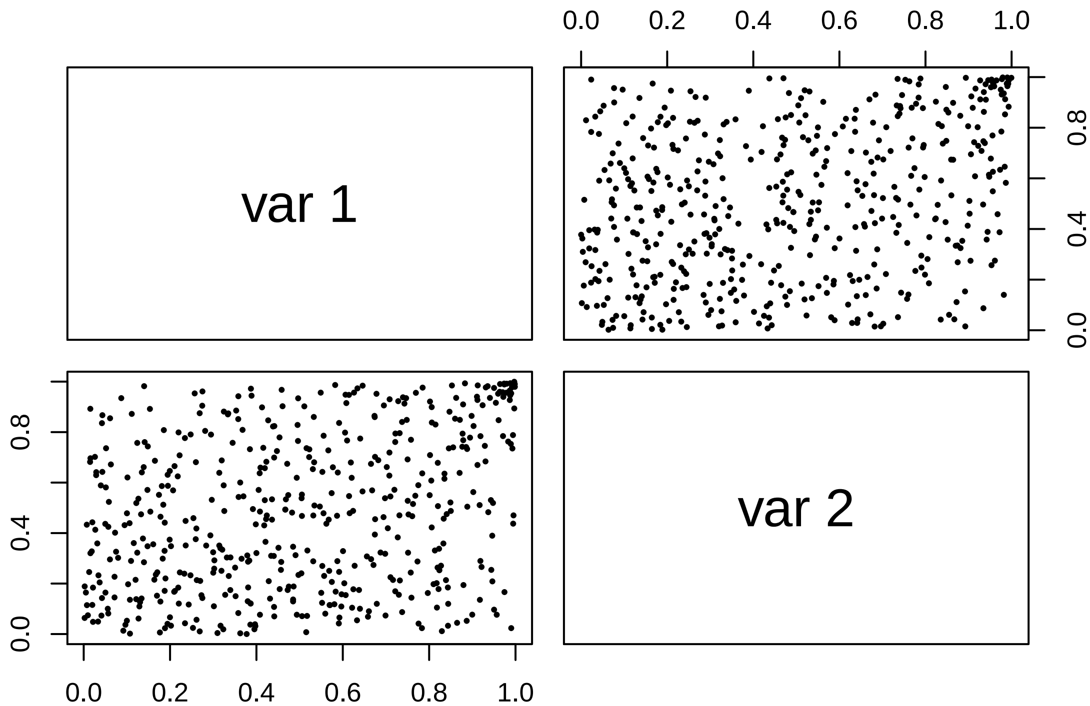
```
pairs(Cop_norm,pch = 16,cex = .5)
```




```r
pairs(Cop_Clay,pch = 16,cex = .5)
```




```r
pairs(Cop_Gumb,pch = 16,cex = .5)
```



Pregunta 11

a. Expliquen cómo utilizarían simulación para estimar $E(L)$.

Partiendo de una permutación $\sigma_\theta = \sigma(1, 2, 3, 4, 5)$ inicial, podemos utilizar el siguiente algoritmo. Para j = 1, . . . , 100: a) Selecciona i $\in (\frac{1}{15}, \frac{2}{15}, \frac{3}{15}, \frac{4}{15}, \frac{5}{15})$ b) Calcula L_j = posición del elemento seleccionado. c) genera la nueva permutación $\sigma(i, X1, X2, X3, X4)$. d) Incrementa j := j + 1 Implementando el algoritmo anterior, obtenemos:

```
sigma0 <- c(1,2,3,4,5) #permutación inicial
n <- 100
L <- I <- NULL
permutaciones <- matrix(rep(0,500),nrow=100,5)
for(j in 1:n){
  i <- sample(1:5,size=1,replace = F,prob = (1:5)/15)
  I[j] <- i
  L[j] <- match(i,sigma0)
  permutaciones[j,] <- c(i,sigma0[-match(i,sigma0)])
  sigma0 <- permutaciones[j,]
  }
Lhat <- mean(L)
Lhat
```

```
## [1] 3.04
```

b. Notemos que $E(N_i) = N_{p_i}$. Entonces es una cantidad fija. Si N = 100, entonces el vector de valores esperados es $(E(N_1), E(N_2), E(N_3), E(N_4), E(N_5)) = 100(\frac{1}{15}, \frac{2}{15}, \frac{3}{15}, \frac{4}{15}, \frac{5}{15})$ = (6.667, 13.333, 20, 26.667, 33.333).

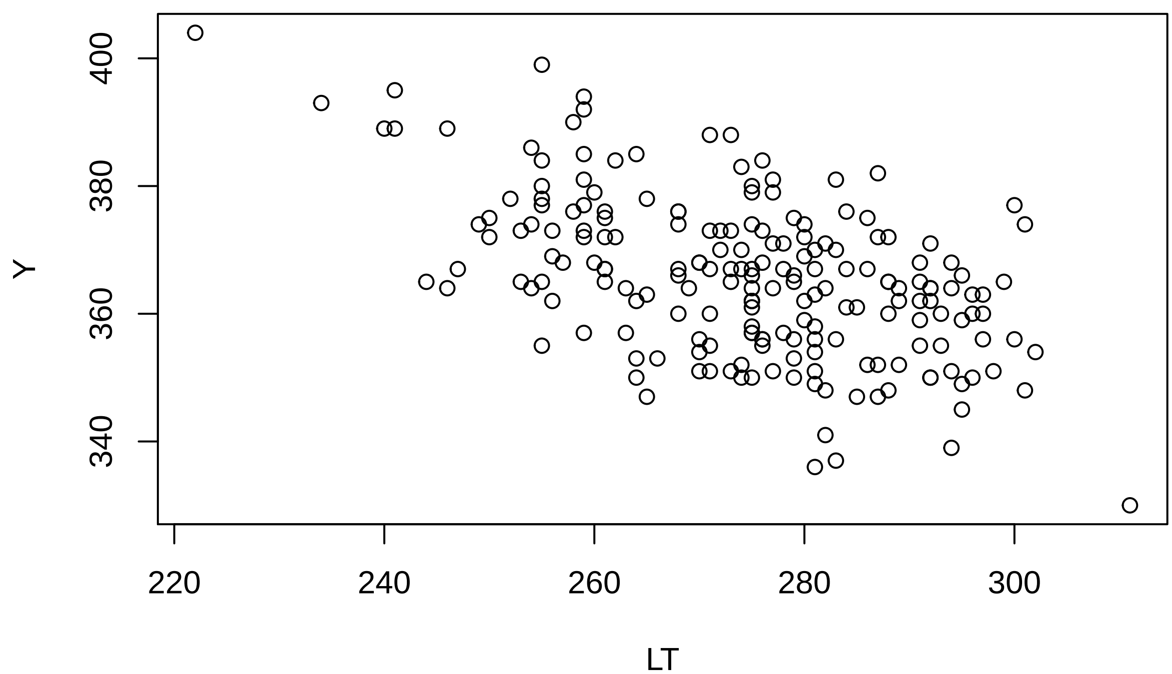
c. Notemos que $E(Y|X) = \sum_{i=1}^{5} ip_i$. Entonces Y|N es el estimador del nÃºmero seleccionado promedio. Ahora bien, Y es la suma promedio de los nÃºmeros seleccionados. Mientras mayor sea la probabilidad de selección de un nÃºmero grande, será más probable que este ocupe la primera posición, haciendo L pequeÃ±a. Entonces parece que la correlación entre Y y L es negativa.

d. Con la consideración de (c), podemos llevar a cabo las simulaciones y generar parejas de puntos (Y, L) para corroborar lo propuesto.

```
simula4 <- function(sigma0=1:5,n=100){
  L <- I <- NULL
  permutaciones <- matrix(rep(0,500),nrow=100,5)
  for(j in 1:n){
    i <- sample(1:5,size=1,replace = F,prob = (1:5)/15)
    I[j] <- i
    L[j] <- match(i,sigma0)
    permutaciones[j,] <- c(i,sigma0[-match(i,sigma0)])
    sigma0 <- permutaciones[j,]
  }
  return(c(LT=sum(L),Y=sum(as.vector(table(I))*(1:5)))) }
datos <- NULL
for(i in 1:200) datos <- rbind(datos,simula4())
```

```
## Warning in as.vector(table(I)) * (1:5): longitud de objeto mayor no es
## múltiplo de la longitud de uno menor
```

```
plot(datos)
```



Entonces, es posible usar Y como variable de control y reducir la varianza, ya que conocemos exactamente $E(Y) = (\frac{100}{15})(\sum_{i=1}^{5} i^2)$= 366.6666667. Entonces el nuevo estimador será $\hat{I} = L +$ Î$E(Y366.67)$, donde ÎË es la pendiente en la línea de regresión estimada y estimamos el nuevo estimador:

```r
summary(lm(LT~Y,data=as.data.frame(datos)))
```

```
##
## Call:
## lm(formula = LT ~ Y, data = as.data.frame(datos))
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -30.800  -9.985   0.337  10.164  33.078
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 514.42335   26.65537  19.299   <2e-16 ***
## Y            -0.65650    0.07286  -9.011   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 12.8 on 198 degrees of freedom
## Multiple R-squared:  0.2908, Adjusted R-squared:  0.2872
## F-statistic: 81.19 on 1 and 198 DF,  p-value: < 2.2e-16
```

```r
betahat <- coefficients(lm(LT~Y,data=as.data.frame(datos)))[2]
datos <- NULL
for(i in 1:1000) datos <- rbind(datos,simula4())
```

```
## Warning in as.vector(table(I)) * (1:5): longitud de objeto mayor no es
## múltiplo de la longitud de uno menor

## Warning in as.vector(table(I)) * (1:5): longitud de objeto mayor no es
## múltiplo de la longitud de uno menor
```

```r
datos <- as.data.frame(datos)
theta <- mean(datos$LT)
tildetheta <-mean(datos$LT-betahat*(datos$Y-366.67))
c(theta,tildetheta)
```

```
## [1] 273.446 273.341
```

```r
#intervalo de confianza para theta
theta +c(-1,1)*qnorm(.975)*sd(datos$LT)/sqrt(1000)
```

```
## [1] 272.5308 274.3612
```

```r
#intervalo de confianza para tildetheta
tildetheta + c(-1,1)*qnorm(.975)*sd(datos$LT - betahat*(datos$Y-366.67))/sqrt(1000)
```

```
## [1] 272.5873 274.0947
```

Pregunta 12

Podemos expresar θ = P(X+Y > 4) = E(I(X+Y > 4)) como un valor esperado. Podemos utilizar el teorema de probabilidad total para encontrar la versión condicionada:

$$
\begin{aligned}
E(I(X+Y>4)) = E(E(I(X+Y>4)|Y)) &= \int_0^\infty P(X+y>4|y)f_y(y)dy \\
&= \int_0^4 P(X>4-y|y)f_y(y)dy \\
&= \int_0^4 P(X>4-y)f_y(y)dy \\
&= E_Y(I(0<y<4)(1-F_x(4-y))) \\
&= E_Y(I(0<y<4)e^{-(4-y)})
\end{aligned}
$$

Entonces podemos estimar $\hat{I}$, con $\hat{I} = \frac{\sum_{i=1}^{n} I(0<y_i<4)e^{-(4-yi)}}{n}$ muestreando de una n exponencial con media 2. Por simetría, el mismo ejercicio es para la otra condición. La que conviene para condicionar es la que tiene menor varianza condicional, ya que la varianza de este estimador depende de la media condicional.

Pregunta 13

Tomando por ejemplo $a = -1$, $X \sim N(-2, 4)$

```
a <- -1
X <- rnorm(100,mean=-2,sd=2)
cor(ifelse(X<a,1,0),X)
```

```
## [1] -0.7571744
```

Utilizando a = 0.3, caso uniforme.

```
k <- 100
a <- 0.3

x <- runif(k)
betaopt <- -lm(ifelse(x<a,1,0)~x)$coeff[2]

x <- runif(1000)
y <- ifelse(x<a,1,0)
vc <- y + betaopt*(x-0.5)
c(mean(y),sd(y))
```

```
## [1] 0.2930000 0.4553662
```

```
c(mean(vc),sd(vc))
```

```
## [1] 0.2949949 0.2794314
```

```
100*(sd(y)-sd(vc))/sd(y)
```

```
## [1] 38.63588
```

Para el caso de X ∼ exponencial, vemos que también alcanzamos reducción de varianza

```
k <- 100
a <- 0.3
x <- rexp(k,1)
betaopt <- -lm(ifelse(x<a,1,0)~x)$coeff[2]
x <- rexp(1000,1)
y <- ifelse(x<a,1,0)
```

```r
vc <- y + betaopt*(x-1)
c(mean(y),sd(y))
```

```
## [1] 0.2740000 0.4462321
```

```r
c(mean(vc),sd(vc))
```

```
## [1] 0.2660250 0.3834769
```

```r
100*(sd(y)-sd(vc))/sd(y)
```

```
## [1] 14.06336
```

Pregunta 14

Para obtener una simulación cruda de p, notemos que podemos escribir $p = P(X \geq 20) = 1 - P(X \leq 20)$ y entonces:

$$\begin{aligned} P(X \geq 20) &= \int_0^1 P(X \geq 20|u)du \\ &= \int_0^1 \sum_{i=0}^{20} e^{-\frac{15}{0.5+u}} \frac{(\frac{15}{0.5+u})^i}{i!} du \\ &= \sum_{i=0}^{20} \frac{1}{i!} \int_0^1 e^{-\frac{15}{0.5+u}} (\frac{15}{0.5+u})^i du \end{aligned}$$

Entonces podemos estimar cada integral g(i) $=\int_0^1 e^{-\frac{15}{0.5+u}} (\frac{15}{0.5+u})^i du$ para cada i = 0, 1, . . . , 20 usando las propias uniformes y evaluando la integral.

```r
g <- NULL
for (i in 0:20){
  u <-runif(10000)
  y <- exp(-15/(0.5+u))*(15/(0.5+u))^i
  g[i+1] <- mean(y)
}
p <- 1 - sum(g*1/factorial(0:20))
p
```

```
## [1] 0.2517457
```

Podemos usar la dependencia entre u y X|u para crear la variable de control.

```r
k <- 100
xu <- NULL
u <- runif(k)
for(i in 1:length(u)) xu[i] <-rpois(1,15/(0.5+u[i]))
betaopt <- -lm(xu ~ u)$coeff[2]
u <- runif(10000)
for(i in 1:length(u)) x[i] <-rpois(1,15/(0.5+u[i]))
vc <- x + betaopt*(u-0.5)
p <- 1- sum(vc<=20)/1000
p
```

```
## [1] -7.005
```

Por último, para el caso de variables antitóticas,

```
g <- NULL
for (i in 0:20){
  u1 <-runif(5000)
  y1 <- exp(-15/(0.5+u))*(15/(0.5+u))^i
  y2 <- exp(-15/(0.5+(1-u)))*(15/(0.5+(1-u)))^i
  g[i+1] <- (mean(y1)+mean(y2))/2
}
p <- 1 - sum(g*1/factorial(0:20))
p
```

```
## [1] 0.2531716
```

Pregunta 15

Por simulación directa:

```
a <- c(1,2,3,1,2)
H <- function(x)min(x[1]+x[4],x[1]+x[3]+x[5],x[2]+x[3]+x[4],x[2]+x[5])
h <- function(x)H(a*x)
E <- NULL
n <- 10000
for(i in 1:n) E[i]<-h(runif(5))
c(mean(E),sd(E)/sqrt(n))
```

```
## [1] 0.928936894 0.003983842
```

Para variable de control, podemos considerar por ejemplo Y = $\min\{X_1 + X_4, X_2 + X_5\}$. Esta variable es particularmente conveniente para los parámetros dados, ya que con alta probabilidad la ruta más corta tendrá longitud igual a Y . Con un poco de cálculos, podemos obtener que E(Y) = 15/16 = 0.9375. Entonces

```
Hc <- function(x)min(x[1]+x[4],x[2]+x[5])
hc <- function(x)Hc(a*x)
u <- matrix(runif(n*5),nrow=n,ncol=5)
Y <- apply(u,1,h)
Yc <- apply(u,1,hc)
a <- cor(Y,Yc)
E <- mean(Y-a*(Yc-15/16))
c(E,sd(Y-a*(Yc-15/16))/sqrt(n))
```

```
## [1] 0.9292903438 0.0005328753
```

Para condicionamiento, definimos $Z_1 = minX_4, X_3 + X_5, Z_2 = minX_5, X_3 + X_4$,entonces $Y_1 = X_1 + Z_1, Y_2 = X_2 + Z_2$ y Y = H(X) se puede escribir como $Y = minY_1, Y_2$ donde condicional a (Z_1, Z_2), (Y_1, Y_2) es uniforme en el rectángulo $[z_1, z_1 + 1] * [z_2, z_2 + 2]$. La esperanza condicional de Y dado (Z_1, Z_2) se puede evaluar exactamente.