Finanças Quantitativas

Lista de exercícios 4

Lucas Emanuel Resck Domingues

Maio de 2020

Problema 3.17 [1]

1.

$$
\begin{aligned}
e^{\sigma^2/2}\,\mathbb{E}\left\{f(Z+\sigma)\right\} &= e^{\sigma^2/2}\int_{-\infty}^{\infty} f(z+\sigma) f_Z(z)dx \\
&= e^{\sigma^2/2}\int_{-\infty}^{\infty} f(w) f_Z(w-\sigma)dw \\
&= e^{\sigma^2/2}\int_{-\infty}^{\infty} f(z) f_Z(z-\sigma)dz \\
&= e^{\sigma^2/2}\int_{-\infty}^{\infty} f(z)\frac{1}{\sqrt{2\pi}}\exp\left\{-\frac{(z-\sigma)^2}{2}\right\}dz \\
&= \int_{-\infty}^{\infty} f(z)\frac{1}{\sqrt{2\pi}}\exp\left\{-z^2/2+z\sigma\right\}dz \\
&= \int_{-\infty}^{\infty} f(z)e^{z\sigma} f_Z(z)dz \\
&= \mathbb{E}\left\{f(Z)e^{Z\sigma}\right\}
\end{aligned}
$$

$$
\begin{aligned}
\mathbb{E}\left\{e^X\right\} &= \mathbb{E}\left\{e^{\mu+\sigma Z}\right\} \\
&= e^{\mu}\,\mathbb{E}\left\{1\cdot e^{\sigma Z}\right\}, \quad f\equiv 1 \\
&= e^{\mu}e^{\sigma^2/2}\,\mathbb{E}\left\{1\right\} \\
&= e^{\mu+\sigma^2/2}
\end{aligned}
$$

2.

$$
\begin{aligned}
\mathbb{E}\left\{e^X h(Y)\right\} &= \int_{\infty}^{\infty}\int_{\infty}^{\infty} e^x h(y) f_{X,Y}(x,y)dxdy \\
&= \int_{\infty}^{\infty}\int_{\infty}^{\infty} e^x h(y+\operatorname{cov}\{X,Y\}) f_{X,Y}(x,y+\operatorname{cov}\{X,Y\})dxdy \\
&\vdots \\
&= \int_{\infty}^{\infty}\int_{\infty}^{\infty} e^x h(y+\operatorname{cov}\{X,Y\}) f_X(x) f_Y(y)dxdy \\
&= \int_{\infty}^{\infty} e^x f_X(x)dx \int_{\infty}^{\infty} h(y+\operatorname{cov}\{X,Y\}) f_Y(y)dy \\
&= \mathbb{E}\left\{e^X\right\}\mathbb{E}\left\{h\left(Y+\operatorname{cov}\{X,Y\}\right)\right\}
\end{aligned}
$$

Problema 3.18 [1]

1. Sabemos que $Y = \log X \sim \mathrm{N}(\mu, \sigma^2)$, logo:

$$
\begin{aligned}
f_X(x) &= f_Y(y)\left|\frac{dy}{dx}\right| \\
&= f_Y(\log x)\left|\frac{d}{dx}(\log x)\right| \\
&= f_Y(\log x)\left|\frac{1}{x}\right| \\
&= \frac{1}{x\sigma\sqrt{2\pi}}\exp\left[-\frac{(\log x-\mu)^2}{2\sigma^2}\right], \quad x>0
\end{aligned}
$$

2, 3. Chamemos ρ o coeficiente de correlação *Pearson* entre $X \sim \text{Lognormal}(0,1)$ e $Y \sim \text{Lognormal}(0,\sigma^2)$. Então queremos calcular:

$$
\rho = \frac{\operatorname{cov}\{X,Y\}}{\sigma_X\sigma_Y} = \frac{\mathbb{E}\{XY\}-\mathbb{E}\{X\}\mathbb{E}\{Y\}}{\sqrt{\operatorname{var}\{X\}\operatorname{var}\{Y\}}}
$$

Vamos calcular cada um desses termos. Se $Z_1 = \log X$ e $Z_2 = \log Y$, então $\mathbf{Z} = (Z_1, Z_2)$ têm distribuição normal multivariada. Vejamos a *moment-generating function* da distribuição normal multivariada:

$$
\begin{aligned}
M_{\mathbf{Z}}(\mathbf{t}) &= \mathbb{E}\left\{e^{\mathbf{t}^T\mathbf{Z}}\right\} \\
&= \mathbb{E}\left\{e^{t_1\log X + t_2\log Y}\right\} \\
&= \mathbb{E}\left\{X^{t_1}Y^{t_2}\right\}
\end{aligned}
$$

Então $M_{\mathbf{Z}}((1,1)) = \mathbb{E}\{XY\}$. Através da expressão da $m.g.f$ da normal multivariada, encontramos:

$$\begin{aligned}
\mathbb{E}\{XY\} &= M_{\mathbf{Z}}((1,1)) \\
&= \exp\left\{\mu^T\mathbf{t} + \frac{1}{2}\mathbf{t}^T\mathbf{\Sigma}\mathbf{t}\right\} \\
&= \exp\left\{\frac{1}{2}\left(1+\sigma^2+2a\right)\right\}, \quad a = \operatorname{cov}\{Z_1, Z_2\}
\end{aligned}$$

De maneira análoga, é possível calcular $\mathbb{E}\{X\} = e^{\frac{1}{2}}$ e $\mathbb{E}\{Y\} = e^{\frac{\sigma^2}{2}}$. Encontramos também $\operatorname{var}\{X\} = (e-1)e$ e $\operatorname{var}\{Y\} = (e^{\sigma^2}-1)e^{\sigma^2}$, a partir dos valores esperados e da definição de variância. Então encontramos ρ:

$$\begin{aligned}
\rho &= \frac{\mathbb{E}\{XY\} - \mathbb{E}\{X\}\,\mathbb{E}\{Y\}}{\sqrt{\operatorname{var}\{X\}\operatorname{var}\{Y\}}} \\
&= \frac{\exp\left\{\frac{1}{2}\left(1+\sigma^2+2a\right)\right\} - e^{\frac{1}{2}}e^{\frac{\sigma^2}{2}}}{\sqrt{(e-1)e(e^{\sigma^2}-1)e^{\sigma^2}}} \\
&= \frac{e^a - 1}{\sqrt{(e-1)\left(e^{\sigma^2}-1\right)}}
\end{aligned}$$

Como $a = \operatorname{cov}\{Z_1, Z_2\}$, encontramos máximo e mínimo para a:

$$\begin{aligned}
-1 \leq \frac{\operatorname{cov}\{Z_1, Z_2\}}{\sigma_{Z_1}\sigma_{Z_2}} \leq 1 \\
-1 \leq \frac{\operatorname{cov}\{Z_1, Z_2\}}{\sigma} \leq 1 \\
-\sigma \leq \operatorname{cov}\{Z_1, Z_2\} \leq \sigma
\end{aligned}$$

Sendo assim, concluímos:

$$\rho_{\min} = \frac{e^{-\sigma} - 1}{\sqrt{(e-1)\left(e^{\sigma^2}-1\right)}}$$

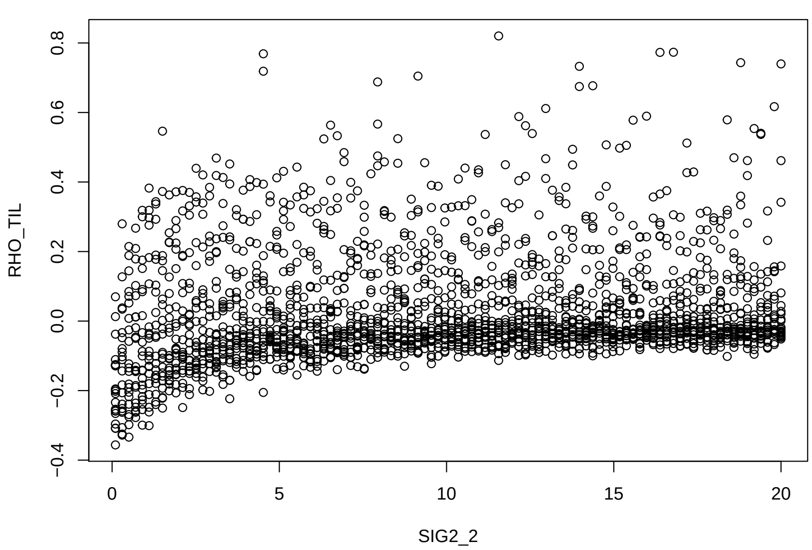
$$\rho_{\max} = \frac{e^{\sigma} - 1}{\sqrt{(e-1)\left(e^{\sigma^2}-1\right)}}$$

4.

$$\begin{aligned}
\lim_{\sigma\to\infty} \rho_{\min} &= \frac{e^{-\sigma} - 1}{\sqrt{(e-1)\left(e^{\sigma^2}-1\right)}} \\
&= 0
\end{aligned}$$

$$
\begin{aligned}
\lim_{\sigma\to\infty}\rho_{\max} &= \frac{e^{\sigma}-1}{\sqrt{(e-1)\left(e^{\sigma^2}-1\right)}}\\
&= \frac{1-1/e^{\sigma}}{\sqrt{(e-1)\left(e^{\sigma^2-2\sigma}-1/e^{2\sigma}\right)}}\\
&= 0
\end{aligned}
$$

Podemos ver que esses resultados condizem com as simulações do **Problema 3.10** [1]:

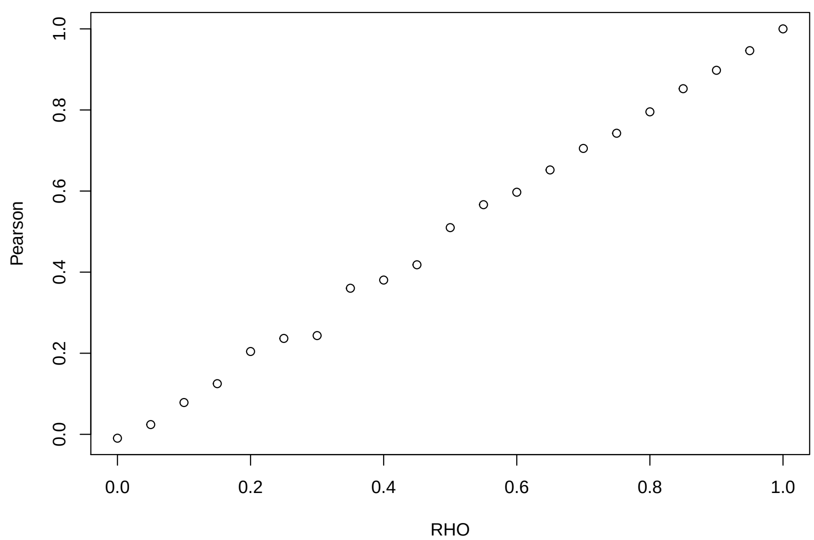


Ou seja, à medida que σ cresce, ρ se aproxima de zero.

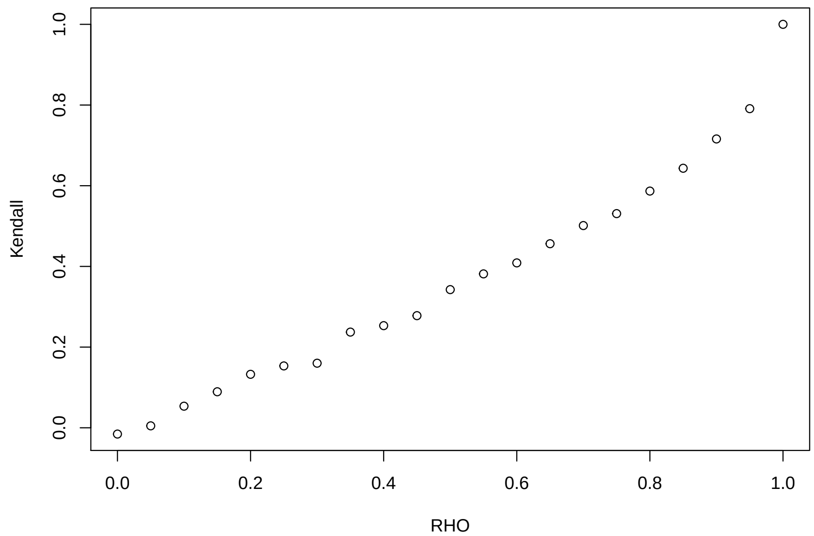
Problema 3.24 [1]

1.
```
suppressMessages(library(copula))
Pearson = list()
Kendall = list()
Spearman = list()
RHO = seq(from = 0, to = 1, by = 0.05)
for(i in 1:length(RHO)) {
```

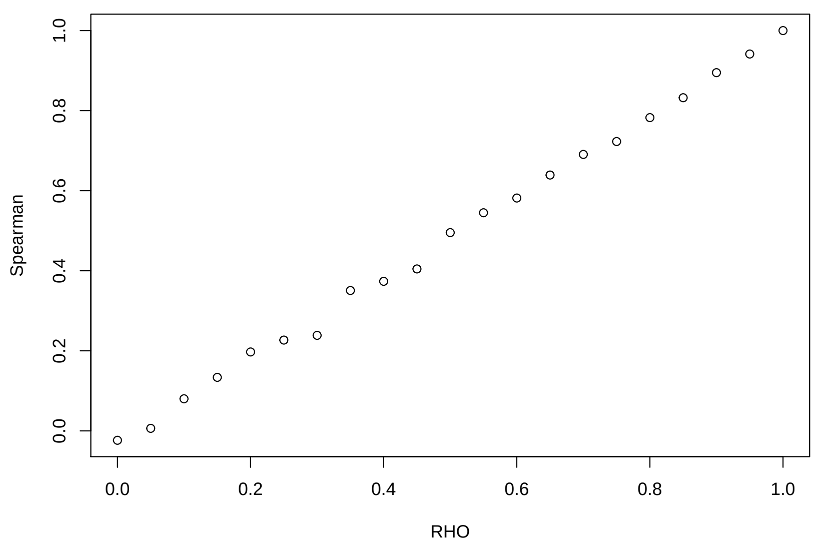
```
  SD = list()
  temp = rCopula(2000, normalCopula(RHO[i], 2))
  SD$x = temp[,1]
  SD$y = temp[,2]
  SD_norm = list()
  SD_norm$x = qnorm(SD$x)
  SD_norm$y = qnorm(SD$y)
  Pearson[[i]] = cor(SD_norm$x, SD_norm$y,
                     method = "pearson")
  Kendall[[i]] = cor(SD_norm$x, SD_norm$y,
                     method = "kendall")
  Spearman[[i]] = cor(SD_norm$x, SD_norm$y,
                      method = "spearman")
}

## Found more than one class "copula" in cache; using the first,
from namespace 'Rsafd'

## Also defined by 'copula'

plot(RHO, Pearson)
```




```
plot(RHO, Kendall)
```



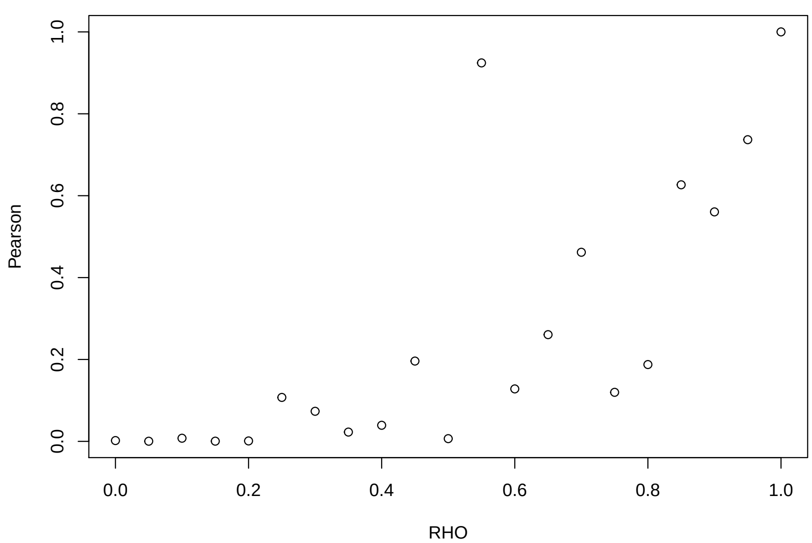

```r
plot(RHO, Spearman)
```



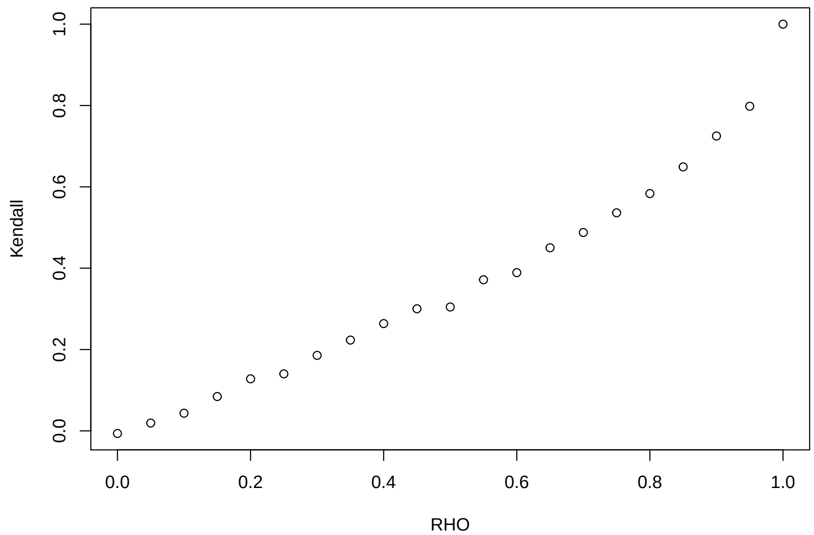
Nos três gráficos encontrados, podemos observar que o coeficiente de correlação cresce praticamente de forma linear com o parâmetro ρ da cópula

Gaussiana. ρ descreve a correlação entre as distribuições marginais. Sendo assim, esperamos que as medidas de correlação, que conhecemos, entre as distribuições marginais amostradas tenham valores próximos a ρ.

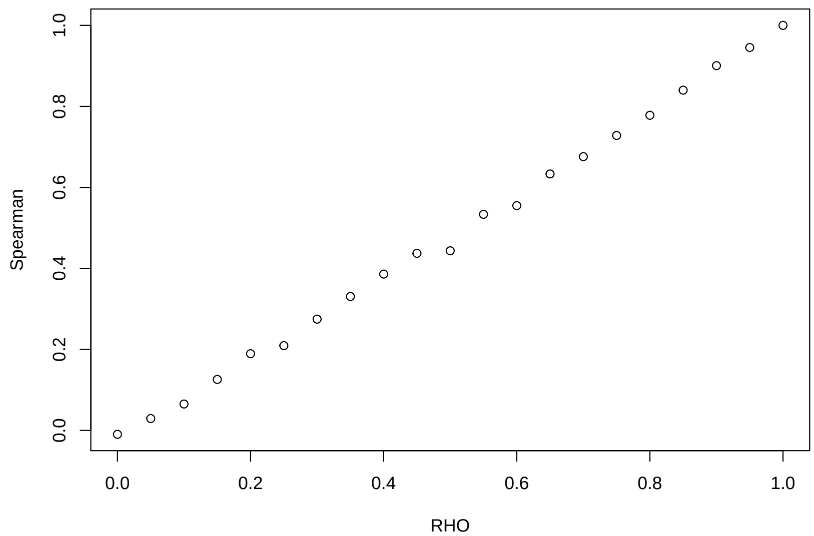
```
Pearson = list()
Kendall = list()
Spearman = list()
RHO = seq(from = 0, to = 1, by = 0.05)
for(i in 1:length(RHO)) {
  SD = list()
  temp = rCopula(2000, normalCopula(RHO[i], 2))
  SD$x = temp[,1]
  SD$y = temp[,2]
  SD_cauchy = list()
  SD_cauchy$x = qcauchy(SD$x)
  SD_cauchy$y = qcauchy(SD$y)
  Pearson[[i]] = cor(SD_cauchy$x, SD_cauchy$y,
                     method = "pearson")
  Kendall[[i]] = cor(SD_cauchy$x, SD_cauchy$y,
                     method = "kendall")
  Spearman[[i]] = cor(SD_cauchy$x, SD_cauchy$y,
                      method = "spearman")
}
plot(RHO, Pearson)
```



```r
plot(RHO, Kendall)
```

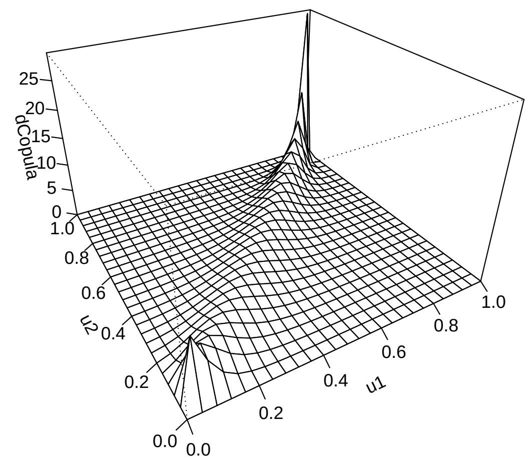


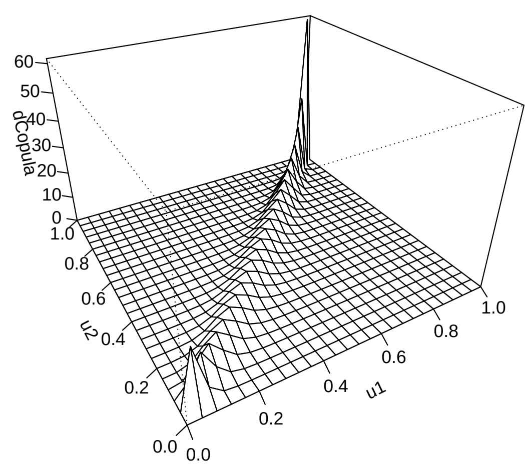

```r
plot(RHO, Spearman)
```

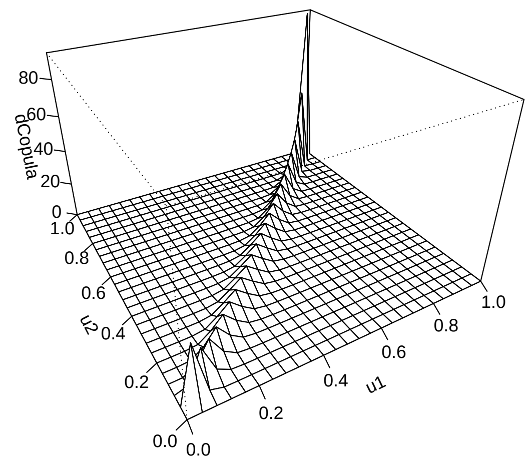


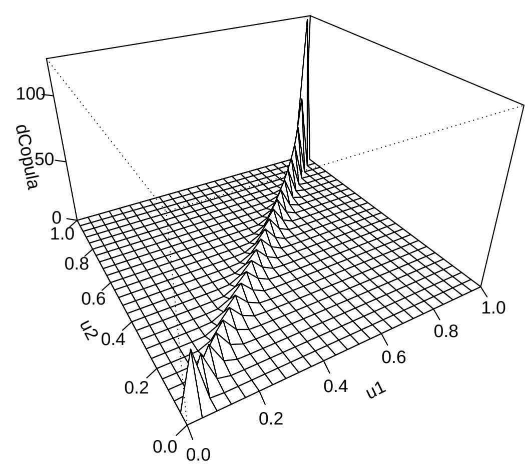
Com exceção do método *Pearson*, os resultados foram parecidos aos anteriores, e as explicações podem ser consideradas as mesmas. No método Pearson, porém, os resultados são muito diferentes. O coeficiente ρ difere bastante do coeficiente de correlação *Pearson*, de modo que o gráfico sequer é crescente. Uma possibilidade é o fato de que o valor esperado de uma distribuição padrão de Cauchy não existe. Sendo assim, não existe a covariância e o coeficiente teórico de correlação *Pearson*, de modo que o coeficiente *Pearson* empírico difere bastante dos outros resultados.

2.
```
Pearson = list()
Kendall = list()
Spearman = list()
BETA = seq(from = 1, to = 20, by = 0.5)
for(i in 1:length(BETA)) {
  SD = list()
  temp = rCopula(2000, gumbelCopula(BETA[i], 2))
  if(BETA[i] %% 5 == 0) {
    persp(gumbelCopula(BETA[i], 2), dCopula)
  }
  SD$x = temp[,1]
  SD$y = temp[,2]
  SD_norm = list()
  SD_norm$x = qnorm(SD$x)
  SD_norm$y = qnorm(SD$y)
  Pearson[[i]] = cor(SD_norm$x, SD_norm$y,
                     method = "pearson")
  Kendall[[i]] = cor(SD_norm$x, SD_norm$y,
                     method = "kendall")
  Spearman[[i]] = cor(SD_norm$x, SD_norm$y,
                      method = "spearman")
}
## parameter at boundary ==> returning indepCopula()
## Found more than one class "copula" in cache; using the first,
from namespace 'Rsafd'
## Also defined by 'copula'
```

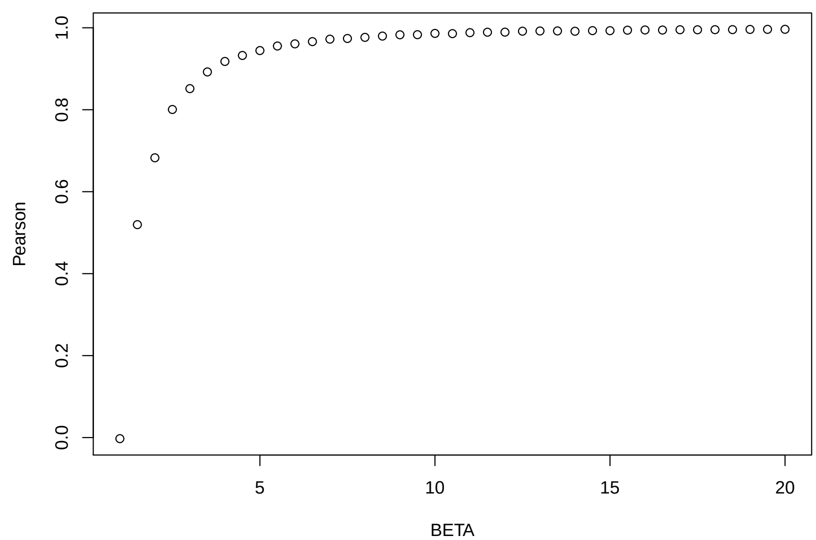

dCopula
0
5
10
15
20
25
u2
0.0
0.2
0.4
0.6
0.8
1.0
u1
0.0
0.2
0.4
0.6
0.8
1.0


dCopula
0
10
20
30
40
50
60
u2
0.0
0.2
0.4
0.6
0.8
1.0
u1
0.0
0.2
0.4
0.6
0.8
1.0

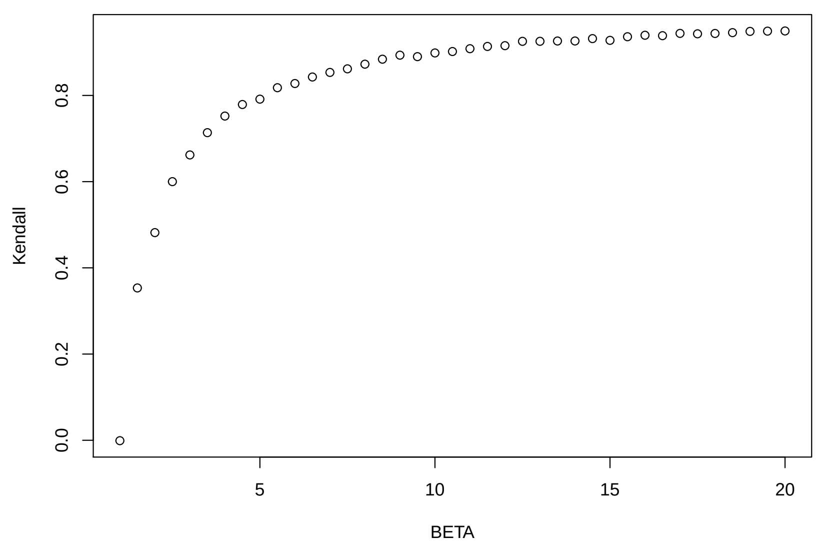




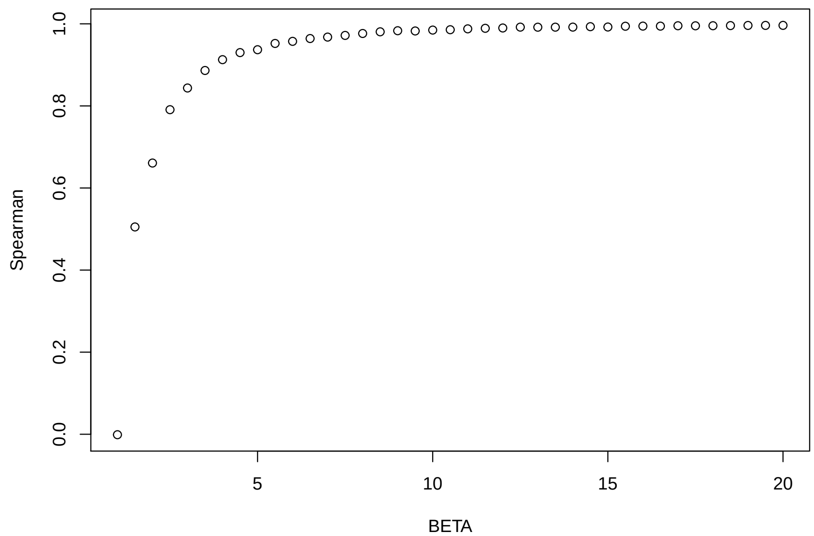

```
plot(BETA, Pearson)
```




```
plot(BETA, Kendall)
```



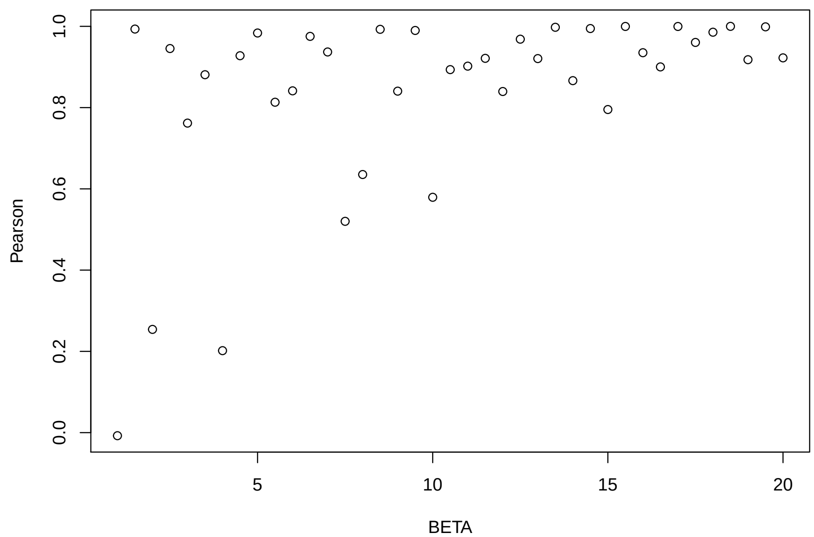
```r
plot(BETA, Spearman)
```



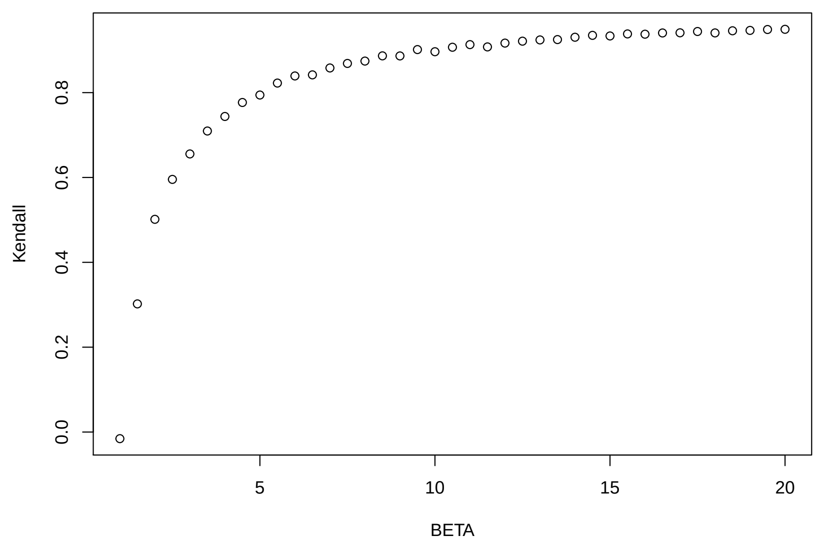
A cópula de Gumbel tem uma propridade interessante: se o valor de β for pequeno, ela se comporta de maneira parecida com uma distribuição uniforme multivariada (ou seja, baixa correlação). Porém, à medida que β cresce, ela obtém o formato que pode ser visto nos gráficos acima. Dessa forma, a correlação entre as marginais cresce muito rápido. Observe que os gráficos correlação vs. β são parecidos entre si.

```r
Pearson = list()
Kendall = list()
Spearman = list()
BETA = seq(from = 1, to = 20, by = 0.5)
for(i in 1:length(BETA)) {
  SD = list()
  temp = rCopula(2000, gumbelCopula(BETA[i], 2))
  SD$x = temp[,1]
  SD$y = temp[,2]
  SD_cauchy = list()
  SD_cauchy$x = qcauchy(SD$x)
  SD_cauchy$y = qcauchy(SD$y)
  Pearson[[i]] = cor(SD_cauchy$x, SD_cauchy$y,
                     method = "pearson")
  Kendall[[i]] = cor(SD_cauchy$x, SD_cauchy$y,
                     method = "kendall")
```

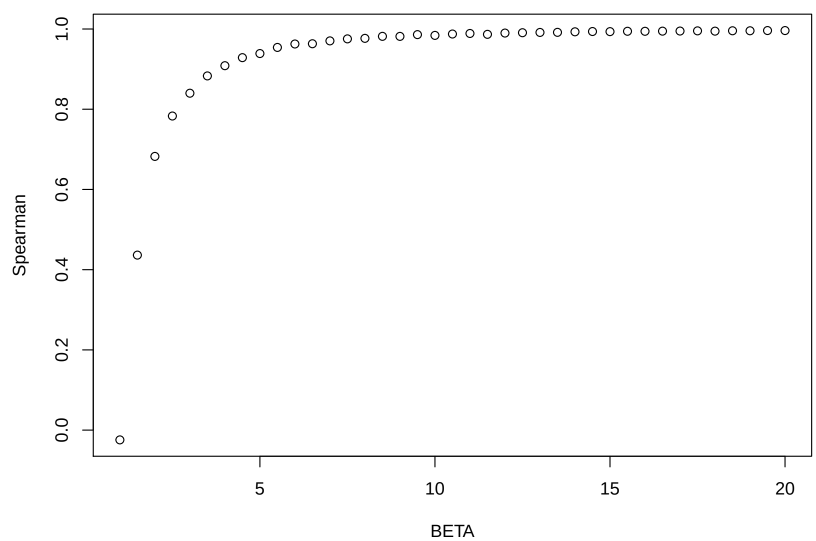
```r
  Spearman[[i]] = cor(SD_cauchy$x, SD_cauchy$y,
                      method = "spearman")
}

## parameter at boundary ==> returning indepCopula()

plot(BETA, Pearson)
```




```r
plot(BETA, Kendall)
```




```r
plot(BETA, Spearman)
```



Com exceção do método *Pearson*, os resultados foram parecidos aos anteriores. Da mesma forma que no exercício 1, uma das possibilidades para o método *Pearson* não funcionar bem é pelo fato de que o valor esperado da Cauchy padrão não existe, como explicado anteriormente.

PCA da Estrutura a Termo da Taxa de Juros (ETTJ)

Importamos e verificamos os dados:

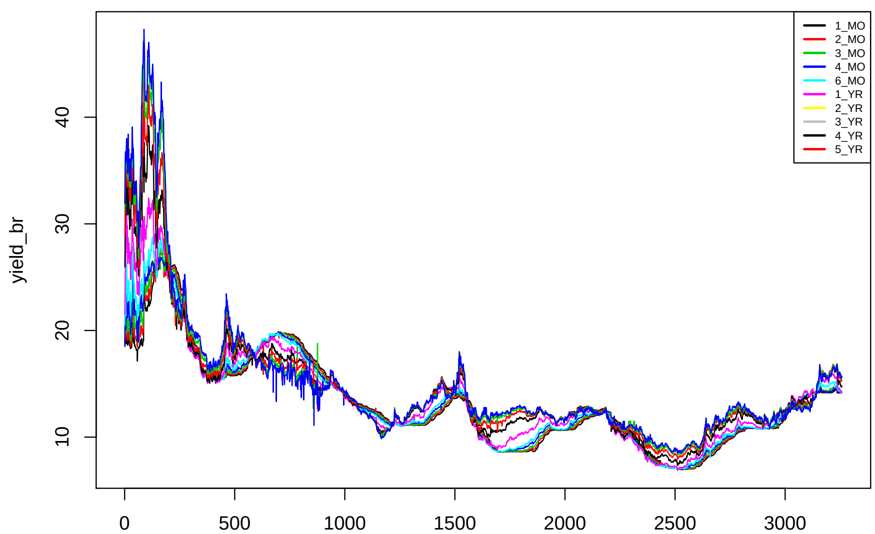
```
load("prices_Swap_PRE_DI.RData")
dim(prices_Swap_PRE_DI)

## [1] 3258   11

head(prices_Swap_PRE_DI)

##         date    4_YR    1_MO    5_YR    2_MO    3_MO    4_MO    6_MO    1_YR
## 1 2002-06-11 31.2753 18.4900 31.9320 18.7200 18.8000 18.7000 20.4900 21.4600
## 2 2002-06-12 33.9903 19.7009 34.6484 19.7090 21.4608 22.3382 23.1222 23.0700
## 3 2002-06-13 35.6196 19.3000 36.2747 20.8400 20.5069 21.0466 22.1055 25.4000
## 4 2002-06-14 36.1113 19.4100 36.7636 20.1900 20.5100 22.4500 22.2600 26.1200
## 5 2002-06-17 35.5415 18.7330 36.1825 19.1634 19.6359 20.1083 21.3597 26.4996
## 6 2002-06-18 35.1315 19.8400 35.7739 19.4000 19.9620 20.7211 22.7835 27.4588
##      2_YR    3_YR
## 1 25.9500 28.9500
## 2 27.7000 30.4100
## 3 30.5200 34.0400
## 4 30.8900 33.9800
## 5 30.7251 33.6000
## 6 31.3700 33.7283
```

Estamos lidando com dados observados uma vez por dia, durante 3258 dias, a partir do dia 11/06/2002. São 10 dimensões, cada uma indicando a taxa de juros para um número específico de meses para a maturidade, neste caso 1, 2, 3, 4, 6, 12, 24, 36, 48, e 60 meses. Vamos organizar o dataframe e visualizar os dados:

```
columns = colnames(prices_Swap_PRE_DI)[c(3, 5, 6, 7, 8,
                                         9, 10, 11, 2, 4)]
yield_br = prices_Swap_PRE_DI[, columns]
head(yield_br)

##      1_MO    2_MO    3_MO    4_MO    6_MO    1_YR    2_YR    3_YR    4_YR
## 1 18.4900 18.7200 18.8000 18.7000 20.4900 21.4600 25.9500 28.9500 31.2753
## 2 19.7009 19.7090 21.4608 22.3382 23.1222 23.0700 27.7000 30.4100 33.9903
## 3 19.3000 20.8400 20.5069 21.0466 22.1055 25.4000 30.5200 34.0400 35.6196
## 4 19.4100 20.1900 20.5100 22.4500 22.2600 26.1200 30.8900 33.9800 36.1113
## 5 18.7330 19.1634 19.6359 20.1083 21.3597 26.4996 30.7251 33.6000 35.5415
```

```
## 6 19.8400 19.4000 19.9620 20.7211 22.7835 27.4588 31.3700 33.7283 35.1315
##      5_YR
## 1 31.9320
## 2 34.6484
## 3 36.2747
## 4 36.7636
## 5 36.1825
## 6 35.7739
```

```
matplot(yield_br, type = "l", lty = 1)
legend("topright",
       legend = colnames(yield_br),
       lwd = 2,
       col=1:length(colnames(yield_br)),
       cex=0.6)
```



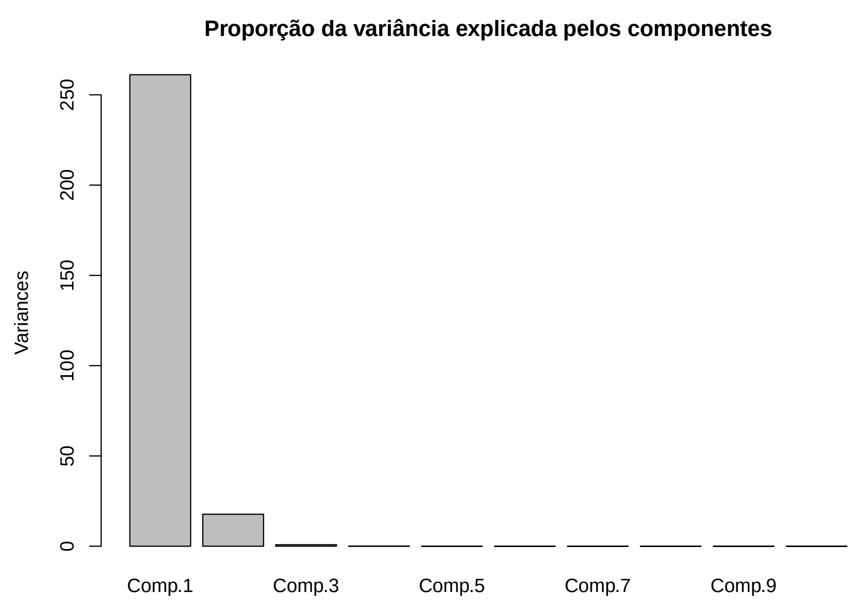
Então podemos realizar a análise de componentes principais (PCA):

```
yield_br.pca = princomp(yield_br)
summary(yield_br.pca)

## Importance of components:
##                          Comp.1     Comp.2      Comp.3       Comp.4
## Standard deviation     16.160223 4.20778556 0.904920443 0.2836770043
```

```
## Proportion of Variance  0.933133 0.06326391 0.002925968 0.0002875392
## Cumulative Proportion   0.933133 0.99639692 0.999322886 0.9996104257
##                              Comp.5       Comp.6       Comp.7       Comp.8
## Standard deviation     0.1959801025 1.437544e-01 1.352633e-01 1.321798e-01
## Proportion of Variance 0.0001372375 7.383988e-05 6.537458e-05 6.242792e-05
## Cumulative Proportion  0.9997476632 9.998215e-01 9.998869e-01 9.999493e-01
##                              Comp.9      Comp.10
## Standard deviation     9.100651e-02 0.0768472113
## Proportion of Variance 2.959333e-05 0.0000211011
## Cumulative Proportion  9.999789e-01 1.0000000000
```

Observe que os três primeiros componentes principais explicam 99,9% da variância, sugerindo um "espaço efetivo" de três dimensões. Qualquer curva pode ser bem aproximada por uma combinação linear dos três primeiros componentes.

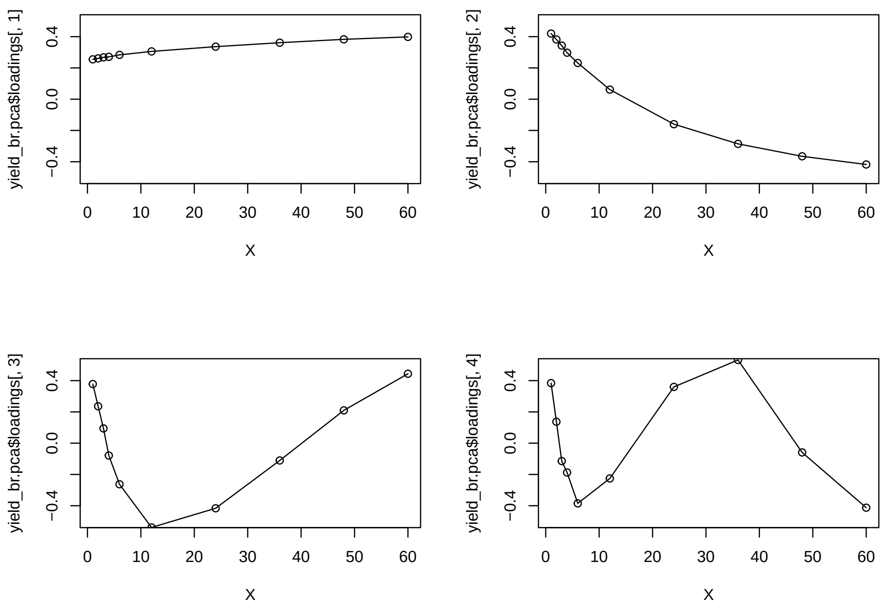
```
plot(yield_br.pca,
     main = "Proporção da variância explicada pelos componentes")
```



Vejamos agora o *plot* dos *loadings*, ou seja, os autovetores:

```
X = c(1, 2, 3, 4, 6, 12, 24, 36, 48, 60)
par(mfrow = c(2, 2))
plot(X, yield_br.pca$loadings[, 1], ylim = c(-0.5, 0.5))
lines(X, yield_br.pca$loadings[, 1], ylim = c(-0.5, 0.5))
```

```r
plot(X, yield_br.pca$loadings[, 2], ylim = c(-0.5, 0.5))
lines(X, yield_br.pca$loadings[, 2], ylim = c(-0.5, 0.5))
plot(X, yield_br.pca$loadings[, 3], ylim = c(-0.5, 0.5))
lines(X, yield_br.pca$loadings[, 3], ylim = c(-0.5, 0.5))
plot(X, yield_br.pca$loadings[, 4], ylim = c(-0.5, 0.5))
lines(X, yield_br.pca$loadings[, 4], ylim = c(-0.5, 0.5))
```



Podemos observar que os gráficos dos primeiros quatro componentes principais resultam como o esperado:

- O primeiro componente é essencilamente constante, de modo que pode ser interpretado como um nível da curva a termos. A intuição é que ele representa os juros médios sobre os vencimentos;
- O segundo componente pode ser interpretado como a tendência da curva a termos, ou seja, se ela cresce ou decresce com o tempo;
- O terceiro representa a curvatura da curva a termos, ou seja, o quanto varia o crescimento ou decrescimento;
- O quarto componente aparentemente não representa nada óbvio.

Referências

[1] Carmona René. *Statistical analysis of financial data in R*. Springer, 2014.