Solutions to Selected Computer Lab Problems and Exercises in Chapter 8 of *Statistics and Data Analysis for Financial Engineering, 2nd ed.* by David Ruppert and David S. Matteson

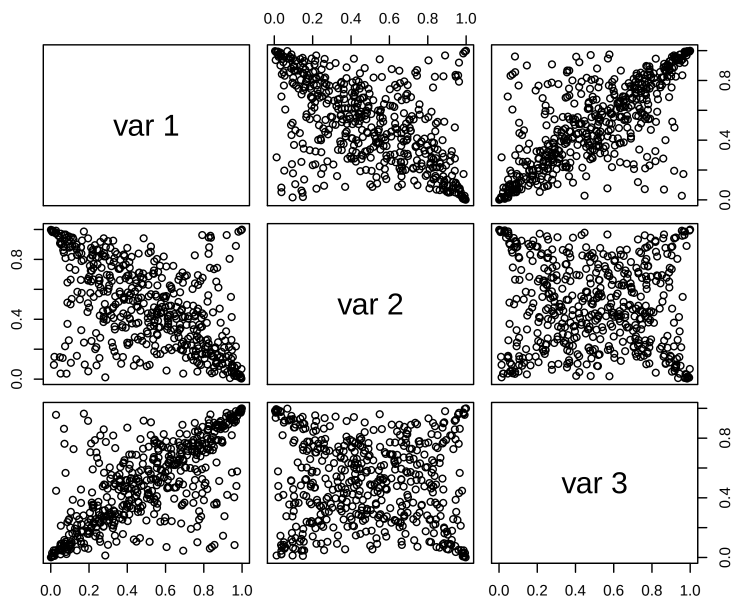
Problem 1. (a) The copula is a t-copula. The correlation matrix is "unstructured" meaning that it is a arbitrary correlation matrix. In fact, the correlation matrix is

$$\begin{pmatrix} 1 & -0.6 & 0.75 \\ -0.6 & 1 & 0 \\ 0.75 & 0 & 1 \end{pmatrix}$$

The degrees of freedom parameter is 1.

(b) The sample size is 500.

Problem 2. (a) The plot is below. Components 2 and 3 would be uniformly scattered over the unit square if they were independent. Clearly, the scatter is not uniform, so they do not appear independent.



(b) The non-uniformity mentioned in (a) is that there are more data in the corners, which shows that extreme values tend to occur together, although because of the zero correlation, a positive extreme value of one component is equally likely to be paired with a positive or negative extreme value of the other component.

(c) The effects of tail dependence is the tendency of extreme values to pair. The negative correlation of components 1 and 2 shows in the concentration of the data along the diagonal from upper left to lower right. Positive extreme values in one component tend to pair with negative extreme values of ther other component.

The positive correlation of components 2 and 3 shows in the concentration of the data along the diagonal from lower left to upper right. Positive extreme values in one component tend to pair with positive extreme values of ther other component.

(d) The output is below and the confidence interval is (0.6603, 0.7484) which does not quite include 0.75. This is not surprising. 0.75 is the correlation between the t-distributed random variables that define the copula and need not be the same as the uniformly-distributed variables in the copula itself.

```
> cor.test(rand_t_cop[,1],rand_t_cop[,3])

Pearson's product-moment correlation

data:  rand_t_cop[, 1] and rand_t_cop[, 3]
t = 22, df = 500, p-value <2e-16
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 0.6603 0.7484
sample estimates:
   cor
0.7071
```

Problem 4. The code and output are below. The estimate `omega` is 0.7018.

```
library(MASS)     #  for fitdistr() and kde2d() functions
library(copula)   #  for copula functions
library(fGarch)   #  for standardized t density
netRtns = read.csv("IBM_SP500_04_14_daily_netRtns.csv", header = T)
ibm = netRtns[,2]
sp500 = netRtns[,3]
est.ibm = as.numeric( fitdistr(ibm,"t")$estimate )
est.sp500 = as.numeric( fitdistr(sp500,"t")$estimate )
est.ibm[2] = est.ibm[2] * sqrt( est.ibm[3] / (est.ibm[3]-2) )
est.sp500[2] = est.sp500[2] * sqrt(est.sp500[3] / (est.sp500[3]-2) )
cor_tau = cor(ibm, sp500, method = "kendall")
omega = sin((pi/2)*cor_tau)
omega

> omega
[1] 0.7018
```

Problem 5. (a) Both fits are by pseudo-likelihood. `ft1` is the parametric approach because the univariate marginal distributions are estimated by fitting t-distributions, and `ft2` is the nonparametric approach because the univariate distributions are estimated by empirical CDFs.

(b) The two estimates of the correlation are 0.7022 and 0.7031. The two estimates of the degrees of freedom are 2.98 and 3.02. Thus, the two estimates of the copula are quite similar with no significant practical difference. Notice also that the two estimates of the correlation are similar to the estimate, 0.7018, in Problem 3 that used Kendall's tau.

```
cop_t_dim2 = tCopula(omega, dim = 2, dispstr = "un", df = 4)
data1 = cbind(pstd(ibm, est.ibm[1], est.ibm[2], est.ibm[3]),
              pstd(sp500, est.sp500[1], est.sp500[2], est.sp500[3]))
n = nrow(netRtns) ; n
data2 = cbind(rank(ibm)/(n+1), rank(sp500)/(n+1))
ft1 = fitCopula(cop_t_dim2, data1, method="ml", start=c(omega,4) )
ft2 = fitCopula(cop_t_dim2, data2, method="ml", start=c(omega,4) )
ft1
ft2

> ft1
fitCopula() estimation based on 'maximum likelihood'
and a sample of size 2516.
      Estimate Std. Error z value Pr(>|z|)
rho.1   0.7022     0.0117    60.1   <2e-16 ***
df      2.9834     0.2693    11.1   <2e-16 ***
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1
The maximized loglikelihood is  967
Optimization converged
Number of loglikelihood evaluations:
function gradient
      46        9
> ft2
fitCopula() estimation based on 'maximum likelihood'
and a sample of size 2516.
      Estimate Std. Error z value Pr(>|z|)
rho.1   0.7031     0.0117    60.3   <2e-16 ***
df      3.0222     0.2785    10.8   <2e-16 ***
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1
The maximized loglikelihood is  965
Optimization converged
Number of loglikelihood evaluations:
function gradient
      39        9
```

Exercise 1. A monotonically strictly decreasing transformation of one variable will change concordant pairs to discordant pairs and vice versa, so it will change the sign but not the magnitude of Kendall's tau.

Since $Y \to 1/Y$ is monotonicity strictly decreasing, Kendall's tau between X and $1/Y$ is -0.55 and Kendall's tau between $1/X$ and $1/Y$ is 0.55.

Exercise 10. By (8.27), $\Omega_{jk} = \sin(0) = 0$, so the copula of (Y_j, Y_k) is the copula of a bivariate Gaussian distribution with correlation matrix equal to the identity matrix. This distribution has independent components so the copula of (Y_j, Y_k) is the independence copula with density identically equal to 1, that is, $c_Y(y_j, y_k) \equiv 1$. Therefore, the components of $\boldsymbol{Y}$ are independent, because, for example, (8.4) implies that

$$f_Y(y_j, y_k) = c_Y\{F_{Y_j}(y_j), F_{Y_k}(y_k)\} f_{Y_j}(y_j) f_{Y_k}(y_k) = f_{Y_j}(y_j) f_{Y_k}(y_k),$$

which shows that Y_j and Y_k are independent.