

# Multivariate t & Copulas

Rob Leonard (robleonard@tamu.edu  
(mailto:robleonard@tamu.edu))

## Question 1: R Lab 7.13.1 Problem 1

Initial setup.

```
library(quantmod); options("getSymbols.warning4.0"=FALSE)
```

```
## Loading required package: xts
```

```
## Warning: package 'xts' was built under R version 4.0.5
```

```
## Loading required package: zoo
```

```
##  
## Attaching package: 'zoo'
```

```
## The following objects are masked from 'package:base':  
##  
##   as.Date, as.Date.numeric
```

```
## Loading required package: TTR
```

```
## Registered S3 method overwritten by 'quantmod':  
##   method             from  
##   as.zoo.data.frame zoo
```

```

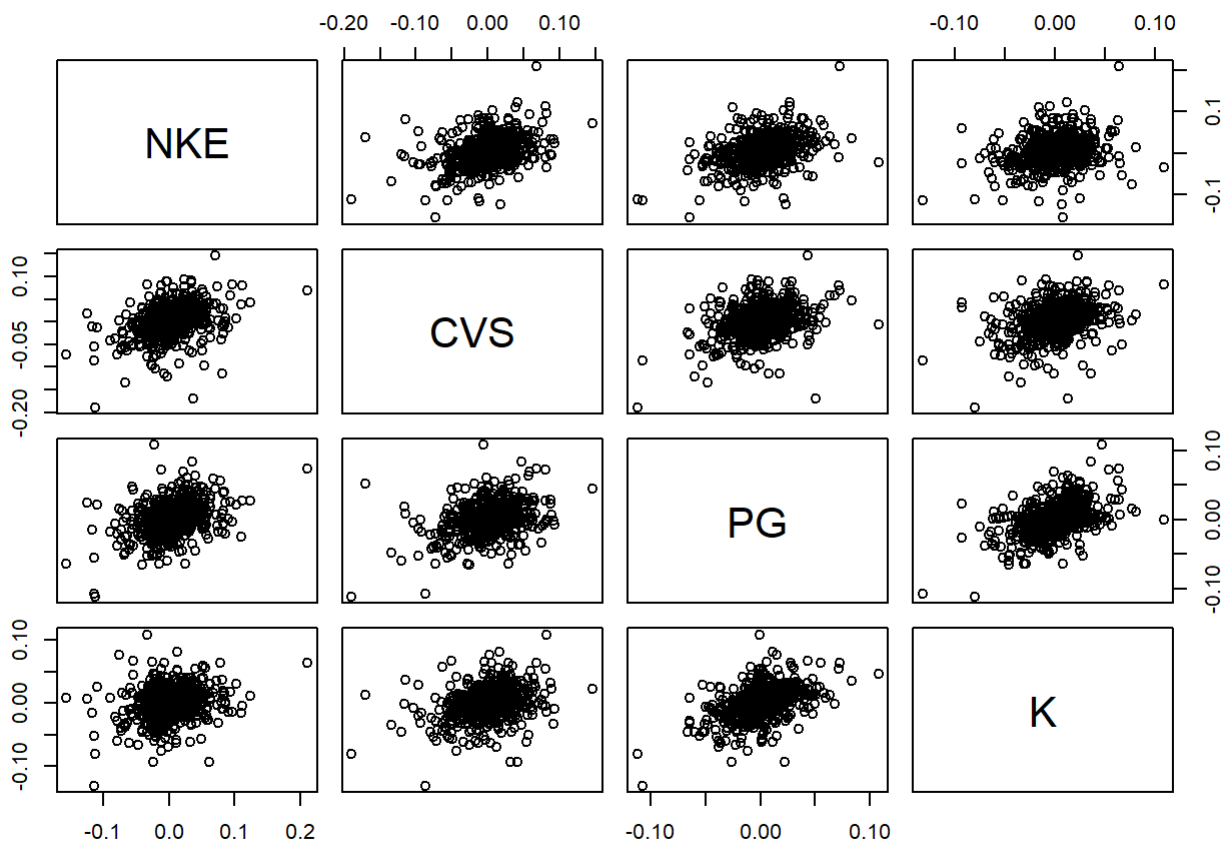
syb          = c("NKE", "CVS", "PG", "K")
d            = length(syb)
rt           = c()
for(i in 1:d) {
  getSymbols(syb[i], from = "2009-01-01", to = "2021-02-01")
  rt = cbind(rt, weeklyReturn(Ad(get(syb[i])), type = "log")[-1])
}
colnames(rt) = syb
rt           = as.matrix(rt) ## remove xts class

# setup skewness and kurtosis functions from HW 02
Sk.fun <- function(x) { ## function to compute skewness from H0 3
  mean((x-mean(x))^3/sd(x)^3)
}
Kur.fun <- function(x){ ## function to compute kurtosis
  mean((x-mean(x))^4/sd(x)^4)
}

```

a) Plot scatterplot and estimate variance.

```
pairs(rt)
```



```

w          = cbind(.5, .3, .2, 0)
var1a      = w %%% cov(rt) %%% t(w)

```

The estimated variance is  $6.710^{-4}$ .

**b) Justify that a multivariate t is a suitable candidate model for  $r_t$ .**

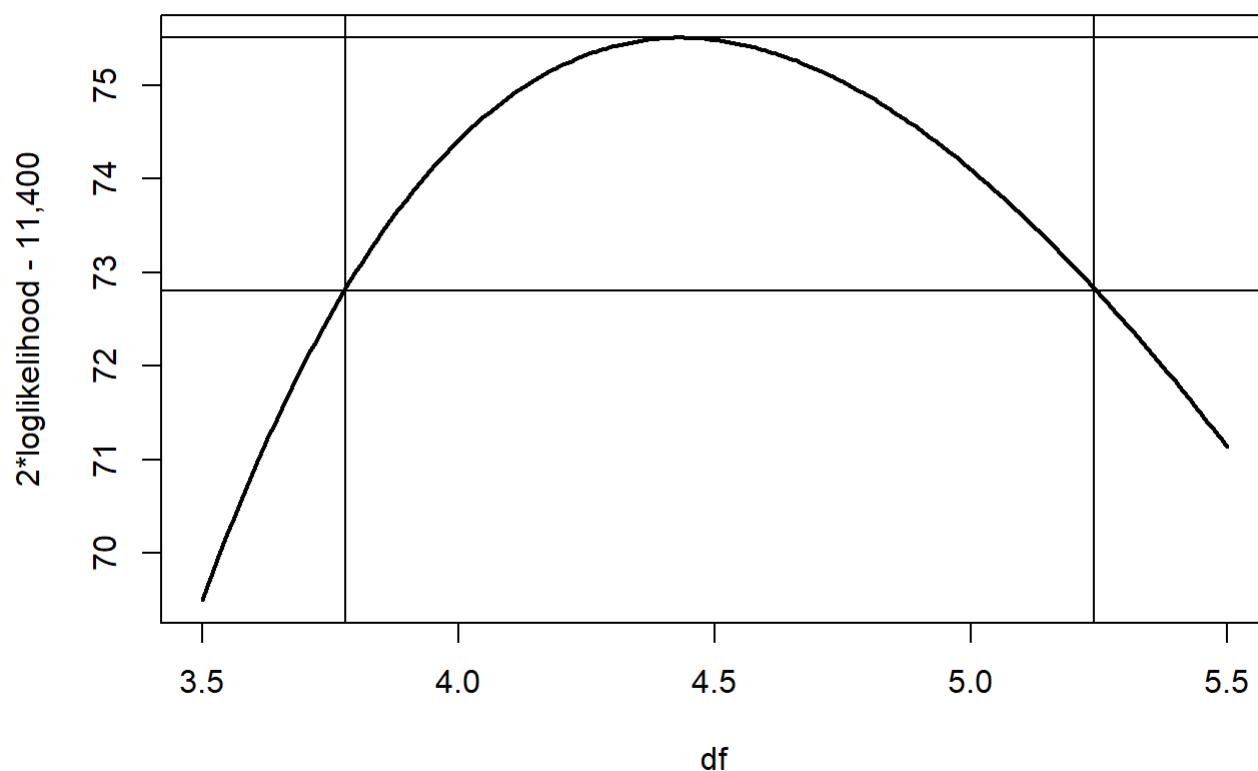
```
library(MASS); options(warn = -1)
stat      = matrix(ncol = dim(rt)[2], nrow=4)
rownames(stat) = c("skew","kurt","nu","se.nu")
colnames(stat) = colnames(rt)
for (i in 1:dim(rt)[2]){
  stat[1,i] = Sk.fun(rt[,i])
  stat[2,i] = Kur.fun(rt[,i])
  fit      = fitdistr(rt[,i], "t")
  stat[3,i] = fit$estimate[3]
  stat[4,i] = fit$sd[3]
}
```

|                 | NKE    | CVS     | PG      | K       |
|-----------------|--------|---------|---------|---------|
| <b>Skewness</b> | 0.056  | -0.6906 | -0.1936 | -0.4235 |
| <b>Kurtosis</b> | 6.3433 | 6.2422  | 5.8685  | 5.6856  |
| <b>nu</b>       | 4.2787 | 4.0744  | 4.1762  | 4.3244  |
| <b>se.nu</b>    | 0.7557 | 0.7174  | 0.7614  | 0.7943  |

**c) Find the MLE of  $v$  and a 90% profile likelihood confidence interval for  $v$ . Also plot profile log likelihood.**

```
library(mnormt)
df      = seq(3.5,5.5, 0.01)
loglik_p = c()
for (i in 1:length(df)) {
  fit = cov.trob(rt, nu = df[i])
  loglik_p[i] = sum(dmt(rt, mean = fit$center, S = fit$cov, df = df[i], log = T))
}
nu      = df[which.max(loglik_p)]

# calc CI boundaries
confintLL = 2*max(loglik_p) - qchisq(.90,1)
lowerind  = which.min(loglik_p[0:100]<= confintLL/2 )
upperind  = which.max(loglik_p[101:201]<= confintLL/2 )+99
# plot
par(mfrow = c(1,1))
plot(df,2*loglik_p-11400,type="l",cex.axis=1,cex.lab=1, ylab="2*loglikelihood - 11,400",lwd=2)
abline(h = 2*max(loglik_p) - qchisq(.90,1)-11400)
abline(h = 2*max(loglik_p)-11400 )
abline(v=df[lowerind])
abline(v=df[upperind])
```



The MLE degrees of freedom is 4.43. The 90% confidence interval is (3.78,5.24).

**d) Calculate the AIC and BIC for multivariate t.**

```
d = dim(rt)[2]; n = dim(rt)[1]
p = d*(d+1)/2 + d + 1
cat(paste(c("d","n","p"), c(d,n,p), sep=": "), sep = ", ")
```

```
## d: 4, n: 630, p: 15
```

```
aic_rt = -2*max(loglik_p) + 2*p
bic_rt = -2*max(loglik_p) + log(n)*p
```

AIC is  $-1.144550710^4$ .

BIC is  $-1.137882110^4$ .

**e) Fit the skewed multivariate t distribution of Azzalini & Capitanio.**

```
library(sn)
```

```
## Loading required package: stats4
```

```
##  
## Attaching package: 'sn'
```

```
## The following object is masked from 'package:stats':  
##  
## sd
```

```
fit_st = mst.mple(y = rt)  
fit_st$dp # estimated nu is greater than 4
```

```
## $beta  
##           NKE           CVS           PG           K  
## [1,] 0.01060784 0.01534077 0.006771056 0.01031285  
##  
## $Omega  
##           NKE           CVS           PG           K  
## NKE 0.0007448777 0.0003282845 0.0001956826 0.0001624949  
## CVS 0.0003282845 0.0008020117 0.0001942983 0.0002474493  
## PG  0.0001956826 0.0001942983 0.0003206882 0.0001834448  
## K   0.0001624949 0.0002474493 0.0001834448 0.0004054260  
##  
## $alpha  
##           NKE           CVS           PG           K  
## -0.04171491 -0.43150045 0.01582096 -0.40581065  
##  
## $nu  
## [1] 4.461239
```

```
dp2cp(fit_st$dp, "st")
```

```
## $beta  
##           NKE           CVS           PG           K  
## [1,] 0.003309943 0.001493388 0.001501732 0.0008158311  
##  
## $var.cov  
##           NKE           CVS           PG           K  
## NKE 0.0012969053 0.0004939914 0.0003162391 0.0002252297  
## CVS 0.0004939914 0.0012619755 0.0002792185 0.0003170175  
## PG  0.0003162391 0.0002792185 0.0005535133 0.0002824691  
## K   0.0002252297 0.0003170175 0.0002824691 0.0006446814  
##  
## $gamma1  
##           NKE           CVS           PG           K  
## -0.4229293 -0.8492520 -0.4691127 -0.8111852  
##  
## $gamma2M  
## gamma2M  
## 112.3345
```

**f) Compute the AIC and BIC of the skewed multivariate-t fit.**

```
p.st = d*(d+1)/2 + d + 1 + d
cat("Number of parameters", paste(p.st))
```

```
## Number of parameters 19
```

```
aic_st = -2*fit_st$logL + 2*p.st
bic_st = -2*fit_st$logL + log(n)*p.st
```

The AIC for the skewed multivariate t distribution is  $-1.144785810^4$ .

The BIC for the skewed multivariate t distribution is  $-1.13633910^4$ .

AIC chooses the skewed multivariate t distribution. BIC chooses the unskewed multivariate t distribution. BIC penalizes extra parameters more than AIC, and we have added 4 skewness parameters. As the AIC values are very similar, I would more rely on the BIC calculation and choose the unskewed multivariate t distribution as the model.

## Question 2

**a) Is a multivariate-t model appropriate for xt?**

```
load("HW04.RData")
```

No, a multivariate t model wouldn't work for all 4 stocks as we would need the estimated degrees of freedom to be comparable to every stock. Here, 3 stocks have degrees of freedom close to 4, but AXP has a value of only 2.64. A multivariate t could work for the other 3 stocks however.

**b) Fit a multivariate t model with the MLE.**

```
x = xt[, c(1,3,4)] # create data set with ACM, CVX and JNJ
dfx      = seq(3, 5, 0.01)
loglik_px = c()
for (i in 1:length(dfx)) {
  fitx = cov.trob(x, nu = dfx[i])
  loglik_px[i] = sum(dmt(x, mean = fitx$center, S = fitx$cov, df = dfx[i], log = T))
}
nux      = dfx[which.max(loglik_px)]
fit2b = cov.trob(x, nu = nux)
varcov2b = fit2b$cov*nux/(nux-2)
```

The MLE degrees of freedom is 4.37.

**c) State the estimated marginal distributions of all series in x.**

The marginal distributions are univariate t distributions with the following parameters (using standard t from HO3 p.40):

|     | mean      | variance  | df   |
|-----|-----------|-----------|------|
| ACM | 0.0470004 | 3.6588712 | 4.37 |

|     | mean      | variance  | df   |
|-----|-----------|-----------|------|
| CVX | 0.0535826 | 1.9180538 | 4.37 |
| JNJ | 0.0673952 | 0.8914852 | 4.37 |

**d) Calculate a likelihood ratio test.**

```
LRT = 2*(loglik_px[138]-loglik_px[101]) # from H0 3 p. 58, 138 corresponds to nu=4.37, 101 to nu=4.0
cat("LRT", LRT, " p-value: ", 1-pchisq(LRT,1))
```

```
## LRT 2.857992 p-value: 0.09092095
```

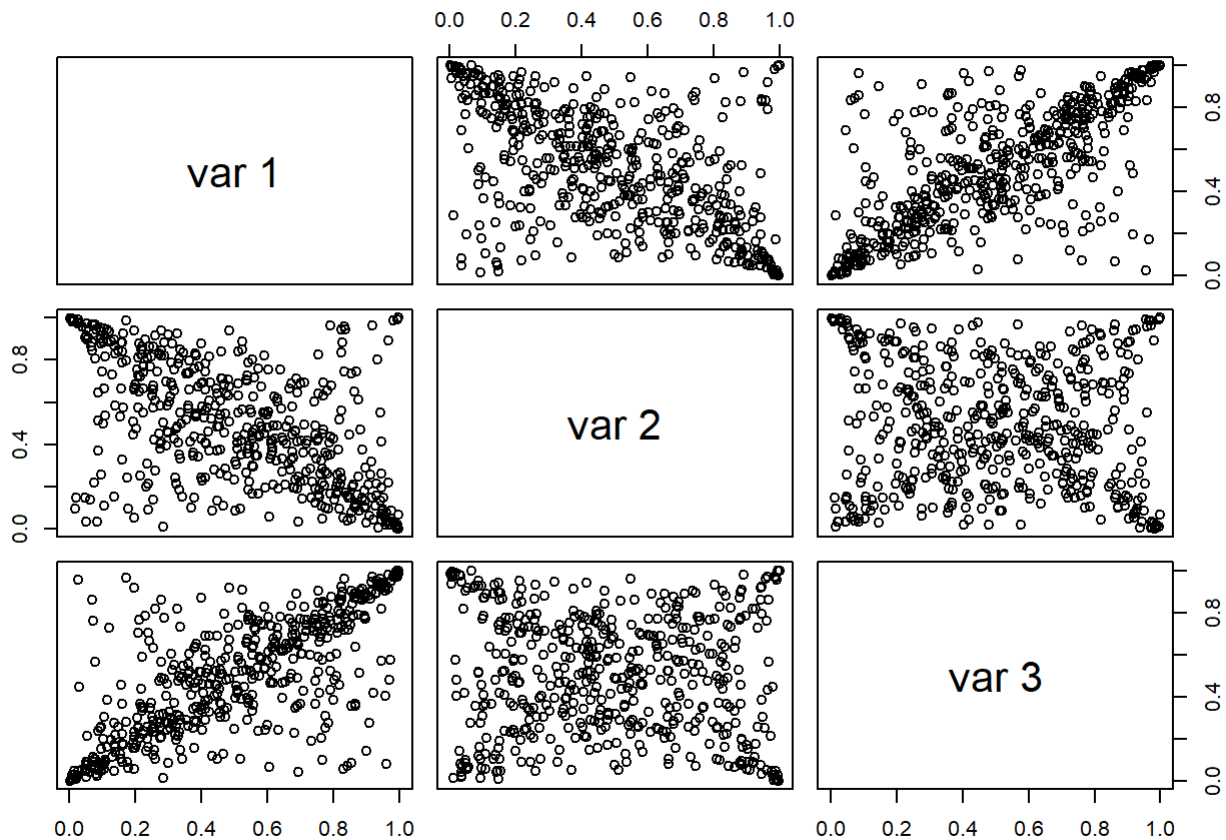
Assuming we use an  $\alpha=0.05$  level test, the p-value of 0.09 is high, so there is not enough evidence to reject the null hypothesis ( $H_0 : \nu = 4$ ). The likelihood ratio test statistic is 2.86.

## Question 7.14

## Question 8.9.1

**Problem 1a) Give the copula family, the correlation matrix, and any other parameters that specify the copula.**

```
library(copula)
cop_t_dim3 = tCopula(dim = 3, param = c(-0.6,0.75,0), dispstr = "un", df = 1)
set.seed(5640)
rand_t_cop = rCopula(n = 500, copula = cop_t_dim3)
pairs(rand_t_cop)
```



```
cor(rand_t_cop)
```

```
##           [,1]      [,2]      [,3]
## [1,]  1.0000000 -0.54999514  0.70707296
## [2,] -0.5499951  1.00000000 -0.06538499
## [3,]  0.7070730 -0.06538499  1.00000000
```

```
corrMatrix = rbind(c(1,-0.6,0.75), c(-0.6,1,0),c(0.75,0,1))
```

The copula family is a t-copula with 1 degree of freedom and 3 dimensions.

The type of positive symmetric positive definite matrix is unstructured ("un").

The correlation matrix is: 1, -0.6, 0.75, -0.6, 1, 0, 0.75, 0, 1

### 1b) What is the sample size?

The sample size is 500.

### Problem 2a) Components 2 and 3 are uncorrelated. Do they appear independent? Why or why not?

They do not appear independent as an independence would mean the scatterplot has no pattern and just fills out a square. There is an X like pattern in the scatterplot, so they are not independent.

### 2b) Do you see signs of tail dependence? If so, where?

Yes, there are signs of tail dependence as there are many points clustered in the corners of the plots, rather than being randomly and more uniformly distributed across the entire plot.



## 2c) What are the effects of dependence upon the plots?

For series 1 and 2, they are negatively correlated, so the tail dependence shows most in the top left and lower right corners. Extreme positive values in one of the series is associated with extreme negative values in the other.

Series 1 and 3 are negatively correlated, so the tail dependence shows most in the lower left and top right corners. Extreme positive values in one series are associated with extreme positive values in the other (and the same for negative values showing as negative in the other).

Series 2 and 3 are uncorrelated, so the tail dependence shows up in all 4 corners.

## 2d) The nonzero correlations in the copula do not have the same values as the corresponding sample correlations. Check the Pearson confidence interval. Explain.

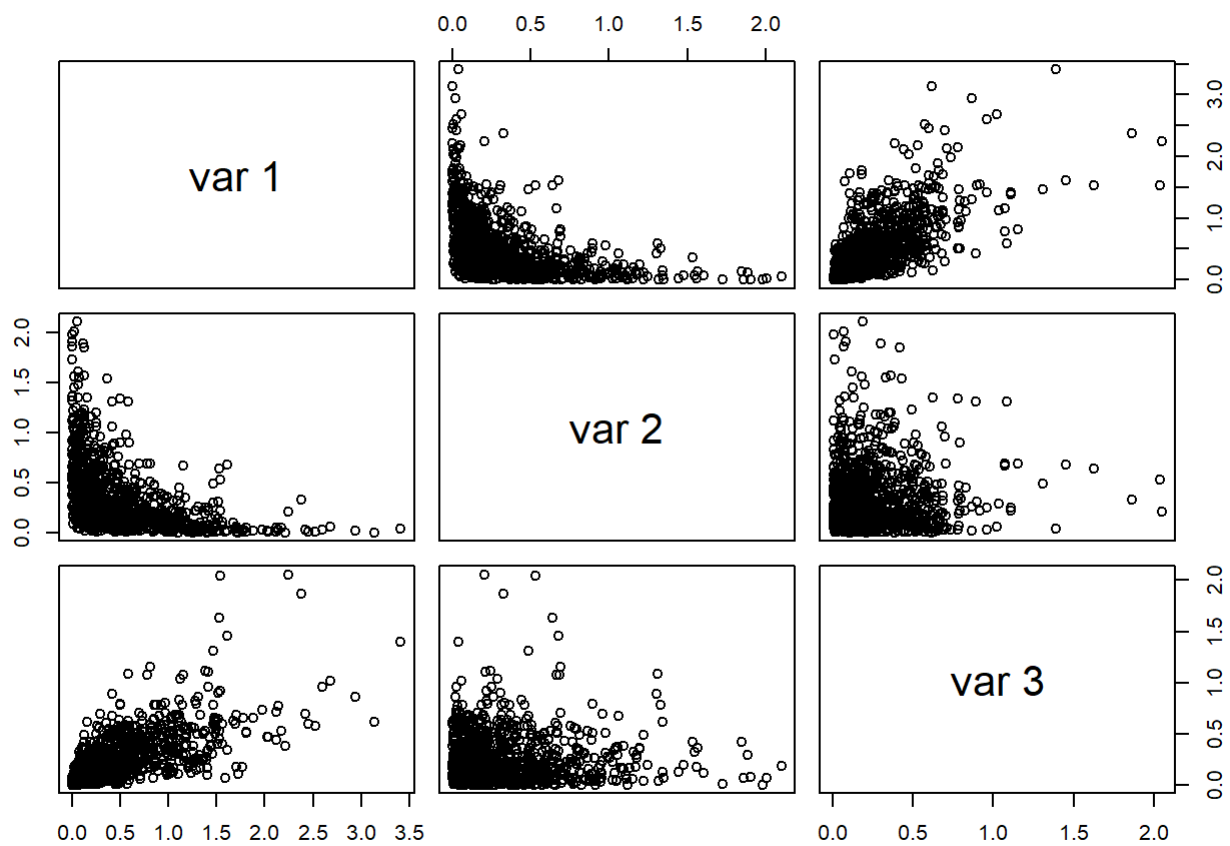
```
cor.test(rand_t_cop[,1],rand_t_cop[,3])
```

```
##
## Pearson's product-moment correlation
##
## data:  rand_t_cop[, 1] and rand_t_cop[, 3]
## t = 22.314, df = 498, p-value < 2.2e-16
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  0.6603249 0.7483624
## sample estimates:
##      cor
## 0.707073
```

The Pearson correlation CI does not include 0.75 in the confidence interval. Our sample correlations are lower than those prescribed in the copula code. The reason for this is that the sample correlation is on uniform random variables inside the copula whereas the 0.75 parameter is for the t-distributed random variables not in the copula.

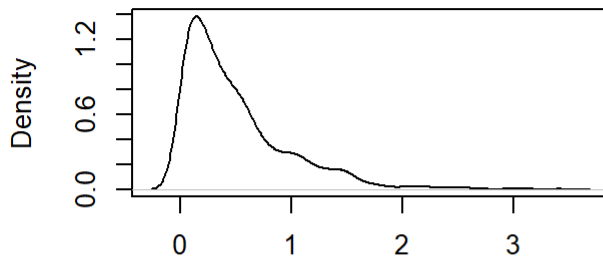
## Problem 3: setup.

```
cop_normal_dim3 = normalCopula(dim = 3, param = c(-0.6,0.75,0), dispstr = "un")
mvdc_normal = mvdc(copula = cop_normal_dim3, margins = rep("exp",3), paramMargins = list(list(rate=2), list(rate=3), list(rate=4)))
set.seed(5640)
rand_mvdc = rMvdc(n = 1000, mvdc = mvdc_normal)
pairs(rand_mvdc)
```



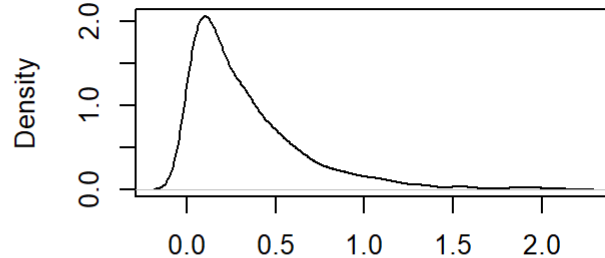
```
par(mfrow = c(2,2))  
for(i in 1:3) plot(density(rand_mvdc[,i]))
```

`density.default(x = rand_mvdc[, i])`



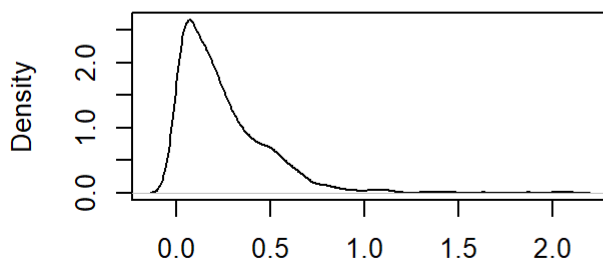
N = 1000 Bandwidth = 0.09159

`density.default(x = rand_mvdc[, i])`



N = 1000 Bandwidth = 0.06263

`density.default(x = rand_mvdc[, i])`



N = 1000 Bandwidth = 0.0464

**3a) What are the marginal distributions of the three components in `rand_mvdc`? What are their expected values?** The marginal distributions are all exponential with rate = 2,3,4, respectively. The expected values are  $1/\text{rate}$ , so  $1/2$ ,  $1/3$  and  $1/4$ .

**3b) Are the second and third components independent? Why or why not?**

Yes, since we are using a gaussian copula and the correlation is set to 0. For gaussian (only) this implies independence.

## Question 8.10

**1) Kendall's tau rank correlation between X and Y is 0.55. Both X and Y are positive. What is Kendall's tau between X and  $1/Y$  ? What is Kendall's tau between  $1/X$  and  $1/Y$  ?**

For the  $1/Y$  transform, this in effect reverses all the ranks and switches the sign. Since the sign of X did not switch, we would end up with a value of -0.55. For both the  $1/X$  and  $1/Y$  transformations, both ranks are now reversed so both signs are switched. But we multiply the signs together and get a net + so the value remains a +0.55.

**2) Suppose that X is Uniform(0,1) and  $Y = X^2$ . Then the Spearman rank correlation and the Kendall's tau between X and Y will both equal 1, but the Pearson correlation between X and Y will be less than 1. Explain why.**

Spearman's rank correlation and Kendall's tau only depend on the ranks of the X's and Y's and not their magnitude. Setting  $Y = X^2$  changes the magnitude but not the ranks. Pearson's correlation depends on the magnitudes of the X's and Y's. The Y values are always less than the X values except when they both are 1, so Pearson's correlation will be less than 1.