Detecting dependence

H. A. Mohtashami-Borzadaran   N. Doodman   M. Amini

Ferdowsi University of Mashhad

Model selection tools for u-data

Now, that the u-scaled data are produced. we are looking for methods to find classes of suitable models, which are

- Dependence structure
  - Kendall plot
  - Dependence measures
- Existence of tail dependence
- Exchangeability of model
- State of radial symmetry

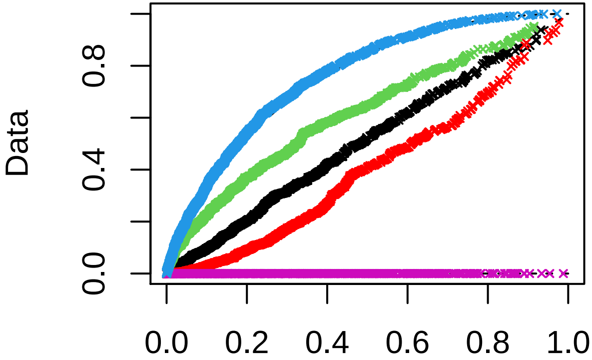
Dependence structure via Kendall plot

In order to see dependence structure of a suitable model for the data, we use the well-known kendall plot:

```
#library(lcopula)
#m1 <- matrix(runif(2000),ncol=2)#Independence
#K.plot(m1, main="Kendall plot")
#m2 <- rCopula(n=1000,claytonCopula(param=-0.5,dim=2))
#K.plot(m2,add=TRUE,col="red") #Negative dependence
#m3 <- rCopula(n=1000,claytonCopula(param=-1,dim=2))
#K.plot(m3,add=TRUE,col="magenta") #Perfect negative dependence
#m4 <- rCopula(n=1000,claytonCopula(param=iTau(claytonCopula(0.3),0.5),dim=2))
#K.plot(m4,add=TRUE,col="green3") #Positive dependence
#m5<- rCopula(n=1000,claytonCopula(param=iTau(claytonCopula(0.3),1),dim=2))
#K.plot(m5,add=TRUE,col="blue") #Perfect positive dependence
```

Kendall plot

The blue line shows Perfect positive dependence, the purple line shows Perfect negative dependence and the black line shows independence.



Kendall plot

So,

- If the data are shown between the black and blue curves, they indicate positive dependence and the strength of positive dependence increases as the data move toward the blue line.
- If the data are shown between the black and purple curves, they indicate negative dependence and the strength of negative dependence increases as the data move toward the purple line.

Dependence measures

The dependence structure of the suitable model for the data can be achieved via measures of central dependence.

Measures of centeral dependence

- Pearson correlation coefficient
- Spearman rank correlation coefficient
- Kendall's Tau coefficient

Dependence structure via Pearson correlation coefficient

Theoretical model based

For any random vector (X, Y), we have:

$$\rho_{X,Y} = \frac{Cov(X, Y)}{\sigma_X \sigma_Y}$$

Empirical estimator

For any set of joint observation (x_i, y_i) we have:

$$r_{xy} = \frac{n \sum x_i y_i - \sum x_i \sum y_i}{\sqrt{n \sum x_i^2 - (\sum x_i)^2} \sqrt{n \sum y_i^2 - (\sum y_i)^2}}$$

Dependence structure via Pearson correlation coefficient

- This measure shows linear correlation.
- The value of this measure is between -1 and $+1$. The value zero indicate non-linear relation between two variables.
- If the observations are normally distributed and their pearson correlation coefficient is zero, then we can say they are independent.
- The values between 0 and $+1$ show positive dependence and the values between -1 and 0 show negative dependence and as the absolute value of the measure increases to 1, the strength of dependence increases.
- If the absolute value of this measure is 1, then the two variables (say X and Y) have the relation $Y = aX + b$, for $a, b \in R$.

Dependence structure via Pearson correlation coefficient

```r
data(danube, package = "lcopula")
cor(danube)
```

```
##           donau       inn
## donau 1.0000000 0.7374098
## inn   0.7374098 1.0000000
```

Dependence structure via Spearman rank correlation coefficient

The Spearman correlation between two variables is equal to the Pearson correlation between the rank values of those two variables; while Pearson's correlation assesses linear relationships, Spearman's correlation assesses monotonic relationships (whether linear or not).

Theoretical model based

$$\rho_{X,Y} = \frac{Cov(rank(X), rank(Y))}{\sigma_{rank(X)}\sigma_{rank(Y)}}.$$

Based on concordance measures:

$$\rho_{X,Y} = 3Q(F, \Pi) = 3Q(C, \Pi) = 12\int_0^1\int_0^1 C(u,v)dudv - 3.$$

Dependence structure via Spearman rank correlation coefficient

The empirical estimator is

$$r_{xy} = \frac{n\sum rank(x_i)rank(y_i) - \sum rank(x_i)\sum rank(y_i)}{\sqrt{n\sum rank(x_i)^2 - (\sum rank(x_i))^2}\,\sqrt{n\sum rank(y_i)^2 - (\sum rank(y_i))^2}}$$

- This measure shows monotone relation of two variable X and Y, that is $Y = g(X)$, where g is a monotone function.
- The value of this measure is between -1 and $+1$. The value zero indicate Y and X have non-monotonic relation.
- The values between 0 and $+1$ show positive dependence and the values between -1 and 0 show negative dependence and as the absolute value of the measure increases to 1, the strength of dependence increases.
- If the absolute value of this measure is 1, then the two variables (say X and Y) have the relation $Y = g(X)$, for the function g.

Dependence structure via Spearman rank correlation coefficient

```r
#rho for data
data(danube, package = "lcopula")
cor(danube,method="spearman")[1,2]
```

```
## [1] 0.7374098
```

```r
#rho for copula models in "Copula" package
gumbel.cop <- gumbelCopula(3)
rho(gumbel.cop)
```

```
## [1] 0.848167
```

Dependence structure via Kendall's Tau coefficient

Theoretical model based

Based on concordance measures:

$$\tau_{X,Y} = Q(F,F) = Q(C,C) = 4\int_0^1\int_0^1 C(u,v)\; dC(u,v) - 1.$$

Empirical estimator

The Kendall's τ coefficient is defined as:

$$\tau = \frac{(number\ \ of\ \ concordant\ \ pairs) - (number\ \ of\ \ discordant\ \ pairs)}{\binom{n}{2}}.$$

Any pair of observations (x_i, y_i) and (x_j, y_j), where $i < j$, are said to be concordant if the ranks for both elements (more precisely, the sort order by x and by y) agree.

Dependence structure via Kendall's Tau coefficient

- the Kendall correlation between two variables will be high when observations have a similar rank between the two variables, and low when observations have a dissimilar rank between the two variables.
- The value of this measure is between -1 and $+1$. The value zero indicate Y and X have non-concordant relation.
- The values between 0 and $+1$ show positive dependence and the values between -1 and 0 show negative dependence and as the absolute value of the measure increases to 1, the strength of dependence increases.

Dependence structure via Kendall's Tau coefficient

```r
#tau for data
data(danube, package = "lcopula")
cor(danube,method="kendall")[1,2]
```

```
## [1] 0.5484731
```

```r
#tau for copula models in "Copula" package
clayton.cop <- claytonCopula(3)
tau(clayton.cop)
```

```
## [1] 0.6
```

```r
#tau for copula models in "VineCopula" package
BiCopPar2Tau(family = 3, par = 3)
```

```
## [1] 0.6
```

```r
#family =3 is the clayton family
```

Existence of tail dependence

The lower tail dependence is defined as

$$\lambda_\ell = \lim_{q \to 0} \mathsf{P}\left(X_2 \le F_2^{\leftarrow}(q) \mid X_1 \le F_1^{\leftarrow}(q)\right),$$

where $F^{\leftarrow}(q) = \inf\{x \in \mathbb{R} : F(x) \ge q\}$, that is, the inverse of the cumulative probability distribution function for q.

The upper tail dependence is defined analogously as

$$\lambda_u = \lim_{q \to 1} \mathsf{P}\left(X_2 > F_2^{\leftarrow}(q) \mid X_1 > F_1^{\leftarrow}(q)\right).$$

Values near zero indicate tail independence and values near 1, show strength of dependence in tails.

Existence of tail dependence (example 1)

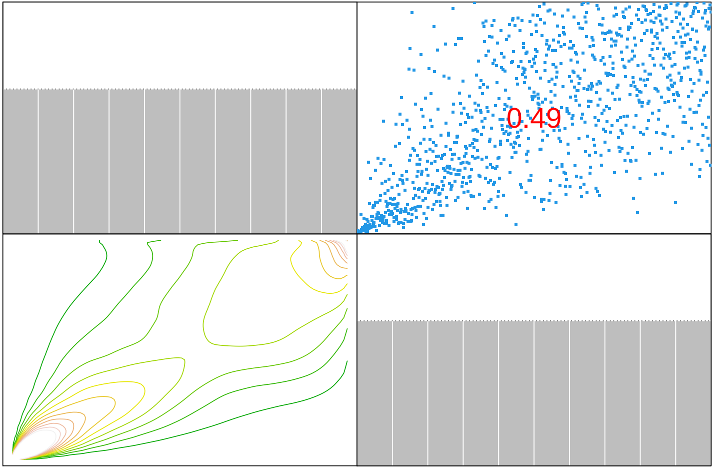
To see the existence of lower/upper tail dependence in the data we have to use the scatter plot & contour plot:

Example: Clayton copula has lower tail dependence.

```
#options(warn=-1) #warnings turned off
#library(copula)
#m1<- rCopula(n=1000,claytonCopula(
#  param=iTau(claytonCopula(0.3),0.5),dim=2))
#library(VineCopula)
#library(kdecopula)
#udata<-pobs(m1)
#pairs.copuladata(udata, method="kendall",
#                 margins = "unif", col=4, cex=4)
#options(warn=0) #warnings turned on
```

Existence of tail dependence (example 1)

Example: Clayton copula has lower tail dependence and the data are more concentrated in the bottom left corner of plots.



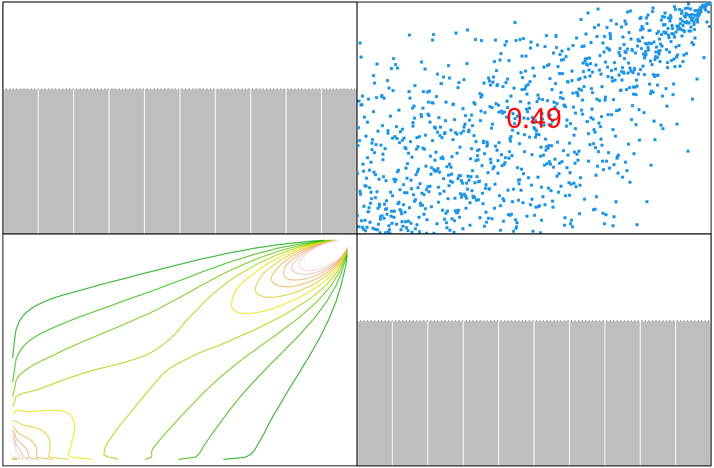
Existence of tail dependence (example 2)

Example: Joe copula has upper tail dependence.

```
#options(warn=-1) #warnings turned off
#library(copula)
#m2<- rCopula(n=1000,joeCopula(
#  param=iTau(joeCopula(2),0.5),dim=2))
#library(VineCopula)
#library(kdecopula)
#udata<-pobs(m2)
#pairs.copuladata(udata, method="kendall",
#                  margins = "unif", col=4, cex=4)
#options(warn=0) #warnings turned on
```

Existence of tail dependence (example 2)

Example: Joe copula has upper tail dependence and the data are more concentrated in the top right corner of plots.



Empirical estimator

There are estimators of lower and upper tail dependence but they are not reliable and can be misleading in their estimate!

- Schmidt, R., & Stadtmueller, U. (2006). Non-parametric estimation of tail dependence. Scandinavian Journal of Statistics, 33(2), 307-335.
- Kiriliouk, A., Segers, J., & Tafakori, L. (2018). An estimator of the stable tail dependence function based on the empirical beta copula. Extremes, 21(4), 581-600.

Test for tail independence

This is the tail dependence test described in Reiss and Thomas (2007) section 13.3.

$$H_0 : (X, Y) \; tail-dependent \quad \text{vs} \quad H_1 : (X, Y) \; tail-independent$$

So, for example, if a significance level alpha $= 0.01$ test is desired, then the null huypothesis (dependence) is rejected for values of the statistic with p-values less than 0.01.

Reference

Reiss, R.-D. and Thomas, M. (2007) Statistical Analysis of Extreme Values: with applications to insurance, finance, hydrology and other fields. Birkh"auser, 530pp., 3rd edition.

Test for tail independence (example 1)

Example: Joe copula has upper tail dependence.

```
library(copula)
m2<- rCopula(n=1000,joeCopula(
  param=iTau(joeCopula(2),0.5),dim=2))
library(extRemes)
taildep.test(m2)
```

```
##
##  Reiss-Thomas (13.35)
##
## data:  m2
## statistic = Inf, threshold = -0.5, m = 223.0, n = 1000.0
## rate (%) = 22.3, p-value = 1
## alternative hypothesis: less
```

Test for tail independence (example 2)

Example: Normal copula has no dependence in the tails.

```
library(copula)
m2<- rCopula(n=1000,normalCopula(0.5,dim=2))
library(extRemes)
taildep.test(m2)
```

```
##
##  Reiss-Thomas (13.35)
##
## data:  m2
## statistic = -4.3407, threshold = -0.5, m = 190.0, n = 1
## exceedance rate (%) = 19.0, p-value = 7.103e-06
## alternative hypothesis: less
```

Exchangeability of model

A copula C is exchangeable if

$$C(u,v) = C(v,u) \qquad \textit{for all } u,v \in (0,1).$$

The hypothesis

$$H_0 : (X,Y) \;\; \textit{Exchangeable} \quad vs \quad H_1 : (X,Y) \;\; \textit{Non-exchangeable}$$

Exchangeability of model

This test is for assessing the exchangeability of the underlying bivariate copula based on the empirical copula. The test statistics are defined in the two references. Approximate p-values for the test statistics are obtained by means of a multiplier technique if there are no ties in the component series of the bivariate data, or by means of an appropriate bootstrap otherwise.

- Genest, C., G. Nešlehová, J. and Quessy, J.-F. (2012). Tests of symmetry for bivariate copulas. Annals of the Institute of Statistical Mathematics 64, 811–834.
- Kojadinovic, I. (2017). Some copula inference procedures adapted to the presence of ties. Computational Statistics and Data Analysis 112, 24–41, http://arxiv.org/abs/1609.05519.

Exchangeability of model (example 1)

Example: Joe copula is an exchangable model.

```r
options(warn=-1) #warnings turned off
library(copula)
m2<- rCopula(n=1000,joeCopula(
  param=iTau(joeCopula(2),0.5),dim=2))
exchTest(m2)
```

```
##
##  Test of exchangeability for bivariate copulas with argument 'm' set to
##  0
##
## data:  m2
## statistic = 0.013394, p-value = 0.8966
```

```r
options(warn=0) #warnings turned on
```

Exchangeability of model (example 2)

Example: khoudraji copula can construct non-exchangable model.

```
options(warn=-1) #warnings turned off
library(copula)
s <- c(0.6,0.95)
copula1 <- gumbelCopula
copula2 <- claytonCopula
param <- function(tau) c(iTau(copula1(), tau),iTau(copula2(), tau))
kho <- khoudrajiCopula(copula1 = copula1(param(0.65)[1])
                        ,copula2 = copula2(param(0.65)[2]),shapes=s)
U <- rCopula(1000,copula=kho)
exchTest(U)
```

```
##
##  Test of exchangeability for bivariate copulas with argument 'm' set to
##  0
##
## data:  U
## statistic = 0.14171, p-value = 0.0004995
```

```
options(warn=0) #warnings turned on
```

State of radial symmetry

A copula C is radially symmetric if and only if $C(u,v) = \bar{C}(u,v)$ for all $u, v \in (0,1)$, where $\bar{C}(u,v)$ is the so-called survival copula defined as

$$\hat{C}(u,v) = u + v - 1 + C(1-u, 1-v).$$

$$H_0 : C(u,v) = \hat{C}(u,v) \qquad vs \qquad H_1 : C(u,v) \neq \hat{C}(u,v)$$

There is relation between tail dependence measures and radial symmetry.

$$\lambda_l = lim_{q \to 0} \frac{C(q,q)}{q} \qquad \text{and} \qquad \lambda_u = lim_{q \to 0} \frac{\hat{C}(q,q)}{q}.$$

The equality of C and $\bar{C}$ states that $\lambda_l = \lambda_u$ for the set of observations!.

State of radial symmetry

This test is for assessing the radial symmetry of the underlying multivariate copula based on the empirical copula. The test statistic is a multivariate extension of the definition adopted in the first reference. An approximate p-value for the test statistic is obtained by means of a appropriate bootstrap which can take the presence of ties in the component series of the data into accont; see the second reference.

- Genest, C. and G. Nešlehová, J. (2014). On tests of radial symmetry for bivariate copulas. Statistical Papers 55, 1107–1119.
- Kojadinovic, I. (2017). Some copula inference procedures adapted to the presence of ties. Computational Statistics and Data Analysis 112, 24–41, http://arxiv.org/abs/1609.05519.

State of radial symmetry (example 1)

Example 1: Frank copula observations are radially symmetric

```r
options(warn=-1) #warnings turned off
frank.cop <- rCopula(200, frankCopula(3))
radSymTest(frank.cop)
```

```
##
##  Test of radial symmetry based on the empirical copula
##
## data:  frank.cop
## statistic = 0.09975, p-value = 0.01449
```

```r
options(warn=0) #warnings turned on
```

State of radial symmetry (example 2)

Example 2: Gumbel copula observations are not radially symmetric

```
options(warn=-1) #warnings turned off
gumbel.cop <- rCopula(500, gumbelCopula(2, dim=3))
radSymTest(gumbel.cop)
```

```
##
##  Test of radial symmetry based on the empirical copula
##
## data:  gumbel.cop
## statistic = 0.25787, p-value = 0.001499
```

```
options(warn=0) #warnings turned on
```

See my homepage

To use the copula course materials, go to the web-page

https://hamb8066.github.io/homepage

and click on teaching section. Choose the "Copula Theory and Applications (Msc)".