Dependence analysis of weights among abalone shells (Application to Lecture 1)

Claudia Czado, Technical University Munich

10 August, 2018

1.The abalone data set and setup

Source and data description

The `abalone` dataset is available from the University of California Irvine (UCI) machine learning repository. Metadata can be obtained from http://archive.ics.uci.edu/ml/datasets/Abalone

- Sex / nominal / – / M, F, and I (infant)
- Length / continuous / mm / Longest shell measurement
- Diameter / continuous / mm / perpendicular to length
- Height / continuous / mm / with meat in shell
- Whole weight / continuous / grams / whole abalone
- Shucked weight / continuous / grams / weight of meat
- Viscera weight / continuous / grams / gut weight (after bleeding)
- Shell weight / continuous / grams / after being dried
- Rings / integer / – / $+1.5$ gives the age in years

Load packages

```r
library(VineCopula) # primary software for vine copulas
library(PivotalR) #contains the data set
library(rafalib) # allows for nicer plots
```

Load data and name columns

The dataset contains 10 variables and 4177 observations. Most of the variables are numeric. The only exception is the `sex` variable. The `rings` variable is slightly different from the other numeric variables because it assumes discrete, integer values.

```
data("abalone")
abalone.cols = c( "sex", "len", "dia", "h", "whole",
                 "shuck", "vis", "shell", "rings")
abalone1=abalone[,-1]
colnames(abalone1)=abalone.cols
sex1=abalone1[,1]
sex.num=rep(0,4177)
sex.num[sex1=="M"]=1
sex.num[sex1=="F"]=0
sex.num[sex1=="I"]=2
abalone1[,1]=sex.num
```

Create datasets for male, female and junvenile separately and remove outlier in height

```
attach(abalone)
abalone.f<-abalone1[sex=="F",-1]
abalone.m<-abalone1[sex=="M",-1]
abalone.i<-abalone1[sex=="I",-1]
detach(abalone)
temp<-max(abalone.f$h)
ind<-(1:length(abalone.f$h))[abalone.f$h==temp]
abalone.fo<-abalone.f[-ind,]
```

2. Exploratory analysis of weight variables for female abalone shells

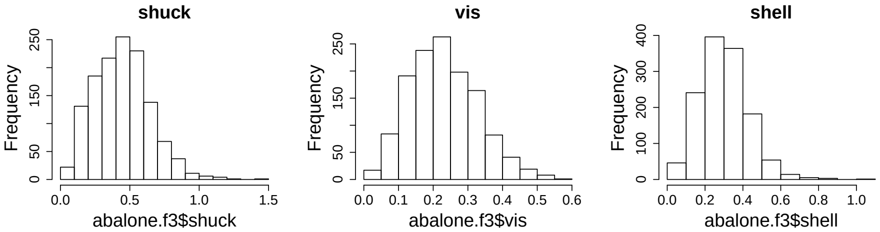
Create weight data set

```
abalone.f3<-abalone.fo[,c("shuck","vis","shell")]
summary(abalone.f3)
```

```
##      shuck             vis             shell       
##  Min.   :0.0310   Min.   :0.0210   Min.   :0.0250  
##  1st Qu.:0.2950   1st Qu.:0.1590   1st Qu.:0.2142  
##  Median :0.4405   Median :0.2240   Median :0.2950  
##  Mean   :0.4463   Mean   :0.2308   Mean   :0.3021  
##  3rd Qu.:0.5734   3rd Qu.:0.2974   3rd Qu.:0.3750  
##  Max.   :1.4880   Max.   :0.5900   Max.   :1.0050  
```

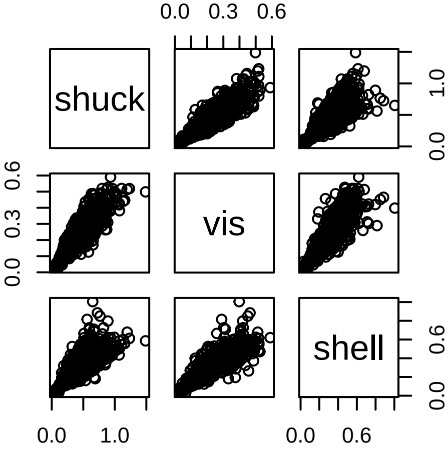
Marginal histograms of weight variables

```
bigpar(1,3)
hist(abalone.f3$shuck, main="shuck")
hist(abalone.f3$vis, main="vis")
hist(abalone.f3$shell, main="shell")
```



Joint EDA for whole, shuck, vis and shell (x-level)

```
pairs(abalone.f3)
```



Assessing strength of dependence

```r
# estimated Pearson correlation
round(cor(abalone.f3),digits=2)
```

```
##       shuck  vis shell
## shuck  1.00 0.89  0.80
## vis    0.89 1.00  0.84
## shell  0.80 0.84  1.00
```

```r
# estimated Kendall's tau
round(cor(abalone.f3, method="kendall"),digits=2)
```

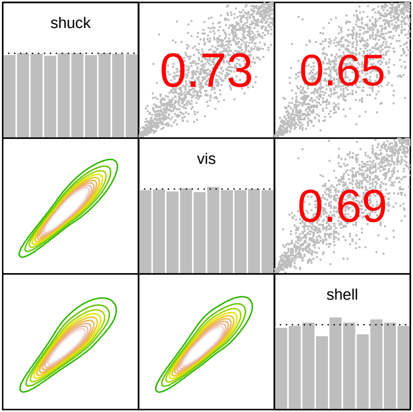
```
##       shuck  vis shell
## shuck  1.00 0.73  0.65
## vis    0.73 1.00  0.69
## shell  0.65 0.69  1.00
```

Use emprical cdfs to transform to copula data

```
n<-nrow(abalone.f3)
fak<-n/(n+1)
temp<-ecdf(abalone.fo$shuck)
u1<-temp(abalone.fo$shuck)*fak
temp<-ecdf(abalone.fo$vis)
u2<-temp(abalone.fo$vis)*fak
temp<-ecdf(abalone.fo$shell)
u3<-temp(abalone.fo$shell)*fak
udata.f3<-cbind(u1,u2,u3)
colnames(udata.f3)<-c("shuck","vis","shell")
udata.f3<-as.copuladata(udata.f3)
```

Empirical normalized contour plots (z-level)

```
pairs(udata.f3)
```



3. Fitting bivariate copulas

Bivariate copula families in VineCopula (Part I)

- 0 = independence copula
- 1 = Gaussian copula
- 2 = Student t copula (t-copula)
- 3 = Clayton copula
- 4 = Gumbel copula
- 5 = Frank copula
- 6 = Joe copula
- 7 = BB1 copula
- 8 = BB6 copula
- 9 = BB7 copula
- 10 = BB8 copula
- 13 = rotated Clayton copula (180 degrees; “survival Clayton”)
- 14 = rotated Gumbel copula (180 degrees; “survival Gumbel”)
- 16 = rotated Joe copula (180 degrees; “survival Joe”)
- 17 = rotated BB1 copula (180 degrees; “survival BB1”)
- 18 = rotated BB6 copula (180 degrees; “survival BB6”)
- 19 = rotated BB7 copula (180 degrees; “survival BB7”)
- 20 = rotated BB8 copula (180 degrees; “survival BB8”)
- 23 = rotated Clayton copula (90 degrees)
- 24 = rotated Gumbel copula (90 degrees)

Bivariate copula families in VineCopula (Part II)

- 26 = rotated Joe copula (90 degrees)
- 27 = rotated BB1 copula (90 degrees)
- 28 = rotated BB6 copula (90 degrees)
- 29 = rotated BB7 copula (90 degrees)
- 30 = rotated BB8 copula (90 degrees)
- 33 = rotated Clayton copula (270 degrees)
- 34 = rotated Gumbel copula (270 degrees)
- 36 = rotated Joe copula (270 degrees)
- 37 = rotated BB1 copula (270 degrees)
- 38 = rotated BB6 copula (270 degrees)
- 39 = rotated BB7 copula (270 degrees)
- 40 = rotated BB8 copula (270 degrees)
- 104 = Tawn type 1 copula
- 114 = rotated Tawn type 1 copula (180 degrees)
- 124 = rotated Tawn type 1 copula (90 degrees)
- 134 = rotated Tawn type 1 copula (270 degrees)
- 204 = Tawn type 2 copula
- 214 = rotated Tawn type 2 copula (180 degrees)
- 224 = rotated Tawn type 2 copula (90 degrees)
- 234 = rotated Tawn type 2 copula (270 degrees)

Bivariate copula estimation for shuck and shell (model parameters)

```
out<-BiCopEstList(udata.f3$shuck,udata.f3$shell,familyset=c(1,2,3,4))
out$models
```

```
## [[1]]
## Bivariate copula: Gaussian (par = 0.84, tau = 0.64)
##
## [[2]]
## Bivariate copula: t (par = 0.85, par2 = 5.45, tau = 0.64)
##
## [[3]]
## Bivariate copula: Clayton (par = 3.27, tau = 0.62)
##
## [[4]]
## Bivariate copula: Gumbel (par = 2.42, tau = 0.59)
##
## [[5]]
## Bivariate copula: Survival Clayton (par = 1.7, tau = 0.46)
##
## [[6]]
## Bivariate copula: Survival Gumbel (par = 2.93, tau = 0.66)
```

Bivariate copula estimation for shuck and shell (model summary)

```
out$summary
```

```
##   family logLik      AIC      BIC
## 1      1 802.79 -1603.58 -1598.40
## 2      2 825.01 -1646.02 -1635.68
## 3      3 894.73 -1787.46 -1782.29
## 4      4 660.07 -1318.14 -1312.96
## 5     13 464.26  -926.52  -921.35
## 6     14 914.87 -1827.74 -1822.57
```

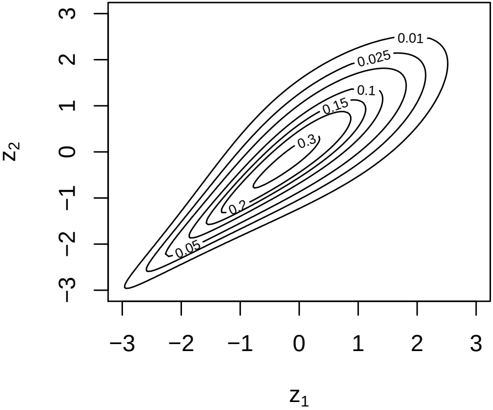
Fitted survival Gumbel model to shuck and shell

```r
est.mle.shuck.shell<-BiCopEst(udata.f3$shuck,
              udata.f3$shell,family=14, method="mle")
summary(est.mle.shuck.shell)
```

```
## Family
## ------
## No:    14
## Name:  Survival Gumbel
##
## Parameter(s)
## ------------
## par:  2.93
##
## Dependence measures
## -------------------
## Kendall's tau:    0.66 (empirical = 0.65, p value < 0.01)
## Upper TD:         0
## Lower TD:         0.73
##
## Fit statistics
## --------------
## logLik:  914.87
## AIC:    -1827.74
## BIC:    -1822.57
```

Fitted survival Gumbel contours for shuck and shell

```
contour(est.mle.shuck.shell)
```



Bivariate copula estimation for shuck and vis (summary)

```
BiCopEstList(udata.f3$shuck,udata.f3$vis,
             familyset=c(1,2,3,4))$summary
```

```
##   family  logLik      AIC      BIC
## 1      1 1109.67 -2217.33 -2212.16
## 2      2 1121.51 -2239.02 -2228.67
## 3      3 1091.75 -2181.49 -2176.32
## 4      4  959.75 -1917.50 -1912.33
## 5     13  718.37 -1434.75 -1429.57
## 6     14 1179.12 -2356.25 -2351.07
```

Bivariate copula estimation for vis and shell (summary)

```
BiCopEstList(udata.f3$vis,udata.f3$shell,
             familyset=c(1,2,3,4))$summary
```

```
##   family  logLik      AIC      BIC
## 1      1  909.10 -1816.19 -1811.02
## 2      2  942.01 -1880.02 -1869.67
## 3      3  961.48 -1920.95 -1915.78
## 4      4  775.46 -1548.92 -1543.75
## 5     13  559.69 -1117.38 -1112.21
## 6     14 1014.30 -2026.59 -2021.42
```

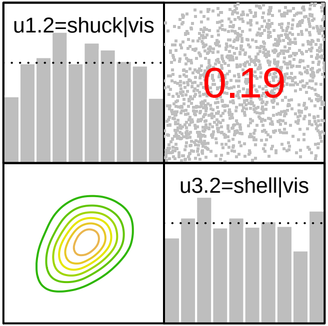
4. Fitting vine with structure c12-c23-c13;2

Fit all survival Gumbel copulas to all bivariate combinations

```
u1<-udata.f3[,"shuck"]
u2<-udata.f3[,"vis"]
u3<-udata.f3[,"shell"]
r12<-BiCopEst(u1,u2,family=14, method="mle")
r13<-BiCopEst(u1,u3,family=14, method="mle")
r23<-BiCopEst(u2,u3,family=14, method="mle")
```

Create pseudo data for exploration for c13;2

```
u1.2<-BiCopHfunc2(u1,u2,family=14,par=r13$par,par2=r13$par2)
u3.2<-BiCopHfunc2(u3,u2,family=14,par=r23$par,par2=r13$par2)
u13.2<-as.copuladata(cbind(u1.2, u3.2))
names(u13.2)<-c("u1.2=shuck|vis","u3.2=shell|vis")
pairs(u13.2,cex.labels = .8)
```



Find copula family for c13;2 using pseudo data

```
BiCopEstList(u1.2,u3.2,familyset=c(1,2,3,4))$summary
```

```
##   family logLik     AIC     BIC
## 1      1  53.82 -105.64 -100.47
## 2      2  54.81 -105.61  -95.26
## 3      3  56.53 -111.06 -105.89
## 4      4  38.90  -75.80  -70.63
## 5     13  30.15  -58.30  -53.12
## 6     14  60.14 -118.27 -113.10
```

Fit survival Gumbel copula to pseudo data (u1.2,u3.2)

```
r13.2<-BiCopEst(u1.2,u3.2,family=14, method="mle")
```

Create R-vine and family matrix for the PCC c12-c23-c13;2

```
Matrix <- c(1, 3, 2,0, 2, 3, 0, 0, 3)
Matrix <- matrix(Matrix, 3, 3)
family <- c(0, 14, 14, 0, 0, 14,0, 0, 0)
family<- matrix(family, 3, 3)
Matrix
```

```
##      [,1] [,2] [,3]
## [1,]    1    0    0
## [2,]    3    2    0
## [3,]    2    3    3
```

```
family
```

```
##      [,1] [,2] [,3]
## [1,]    0    0    0
## [2,]   14    0    0
## [3,]   14   14    0
```

Create parameter matrix for the PCC c12-c23-c13;2

```
par23<-r23$par
par12<-r12$par
par13.2<-r13.2$par
par <- c(0, par13.2, par23,
            0, 0, par12,
            0, 0, 0)
par<- matrix(par, 3, 3)
par2<-matrix(0,3,3)
RVM <- RVineMatrix(Matrix = Matrix, family = family,
                   par = par, par2 = par2,
                   names = c("shuck", "vis", "shell"))
```

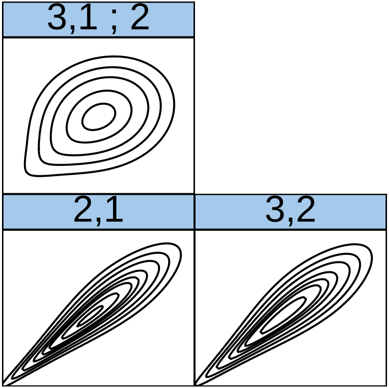
Fit R-vine with specified RVM (c12-c23-c13;2)

```
out13.2fixed<-RVineSeqEst(udata.f3, RVM=RVM, method = "mle")
summary(out13.2fixed)
```

```
## tree   edge | family  cop   par  par2 |  tau   utd   ltd
## ---------------------------------------------------------
##    1    2,1 |     14   SG  3.65  0.00 | 0.73     -  0.79
##         3,2 |     14   SG  3.21  0.00 | 0.69     -  0.76
##    2  3,1;2 |     14   SG  1.23  0.00 | 0.19     -  0.24
## ---
## type: C-vine    logLik: 2253.91    AIC: -4501.82    BIC: -4486.29
## ---
## 1 <-> shuck,   2 <-> vis,   3 <-> shell
```

Fitted normalized contour plots for specified RVM (c12-c23-c13;2)

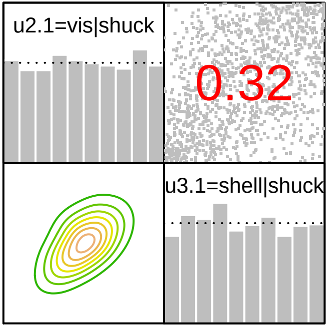
```r
contour(out13.2fixed)
```



5. Fitting vine c13-c12-c23;1

Create pseudo data for exploration for c23;1

```r
u2.1<-BiCopHfunc2(u2,u1,family=14,par=r12$par,par2=r12$par2)
u3.1<-BiCopHfunc2(u3,u1,family=14,par=r13$par,par2=r13$par2)
u23.1<-as.copuladata(cbind(u2.1, u3.1))
names(u23.1)<-c("u2.1=vis|shuck","u3.1=shell|shuck")
pairs(u23.1,cex.labels = .8)
```



Find copula family for c23;1 using pseudo data

```
BiCopEstList(u2.1,u3.1,familyset=c(1,2,3,4))$summary
```

```
##   family logLik     AIC     BIC
## 1      1 151.25 -300.50 -295.33
## 2      2 156.97 -309.93 -299.58
## 3      3 129.62 -257.23 -252.06
## 4      4 125.49 -248.98 -243.81
## 5     13 100.94 -199.88 -194.70
## 6     14 153.48 -304.96 -299.79
```

Fit survival Gumbel copula to pseudo data (u2.1,u3.1)

```
r23.1<-BiCopEst(u2.1,u3.1,family=14, method="mle")
```

Create R-vine and family matrix for the PCC c13-c12-c23;1

```
Matrix <- c(2, 3, 1,0, 1, 3, 0, 0, 3)
Matrix <- matrix(Matrix, 3, 3)
family <- c(0, 14, 14, 0, 0, 14,0, 0, 0)
family<- matrix(family, 3, 3)
Matrix
```

```
##      [,1] [,2] [,3]
## [1,]    2    0    0
## [2,]    3    1    0
## [3,]    1    3    3
```

```
family
```

```
##      [,1] [,2] [,3]
## [1,]    0    0    0
## [2,]   14    0    0
## [3,]   14   14    0
```

Create parameter matrix for the PCC c13-c12-c23;1

```
par13<-r13$par
par12<-r12$par
par23.1<-r23.1$par
par <- c(0, par23.1, par12,
         0, 0, par13,
         0, 0, 0)
par<- matrix(par, 3, 3)
par2<-matrix(0,3,3)
RVM <- RVineMatrix(Matrix = Matrix, family = family,
                   par = par, par2 = par2,
                   names = c("shuck", "vis", "shell"))
```

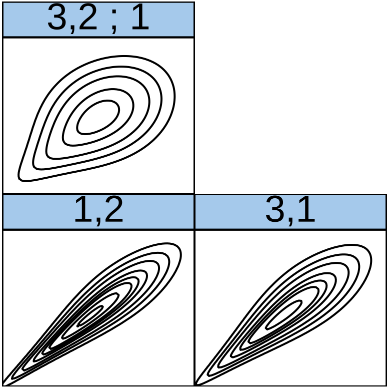
Fit R-vine with specified RVM (c13-c12-c23;1)

```
out23.1fixed<-RVineSeqEst(udata.f3, RVM=RVM, method = "mle")
summary(out23.1fixed)
```

```
## tree   edge | family  cop   par  par2 |  tau   utd   ltd
## --------------------------------------------------------
##    1    1,2 |     14   SG  3.65  0.00 | 0.73     -  0.79
##         3,1 |     14   SG  2.93  0.00 | 0.66     -  0.73
##    2  3,2;1 |     14   SG  1.43  0.00 | 0.30     -  0.38
## ---
## type: C-vine    logLik: 2247.47    AIC: -4488.95    BIC: -4473.42
## ---
## 1 <-> shuck,   2 <-> vis,   3 <-> shell
```

Fitted normalized contour plots for specified RVM (c13-c12-c23;1)

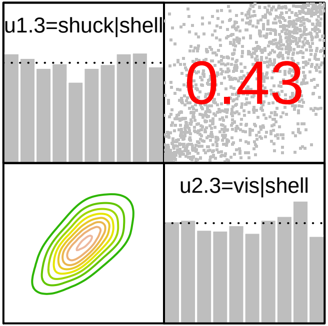
```
contour(out23.1fixed)
```



6. Fitting vine c13-c23-c12;3

Create pseudo data for exploration for c_12;3

```
u1.3<-BiCopHfunc2(u1,u3,family=14,par=r13$par,par2=r13$par2)
u2.3<-BiCopHfunc2(u2,u3,family=14,par=r23$par,par2=r23$par2)
u12.3<-as.copuladata(cbind(u1.3, u2.3))
names(u12.3)<-c("u1.3=shuck|shell","u2.3=vis|shell")
pairs(u12.3,cex.labels=.8)
```



Find copula family for c12;3 using pseudo data (u1.3,u2.3)

```
BiCopEstList(u1.3,u2.3,familyset=c(1,2,3,4))$summary
```

```
##   family logLik     AIC     BIC
## 1      1 318.14 -634.29 -629.11
## 2      2 324.08 -644.17 -633.82
## 3      3 271.15 -540.30 -535.12
## 4      4 284.97 -567.95 -562.77
## 5     13 226.36 -450.72 -445.55
## 6     14 310.30 -618.61 -613.43
```

Fit Student t copula to pseudo data (u1.3,u2.3)

```
r12.3<-BiCopEst(u1.3,u2.3,family=2, method="mle")
```

Create R-vine and family matrix for the PCC c13-c23-c12;3

```
Matrix <- c(1, 2, 3,0, 2, 3, 0, 0, 3)
Matrix <- matrix(Matrix, 3, 3)
family <- c(0, 2, 14,0, 0, 14, 0, 0, 0)
family<- matrix(family, 3, 3)
Matrix
```

```
##      [,1] [,2] [,3]
## [1,]    1    0    0
## [2,]    2    2    0
## [3,]    3    3    3
```

```
family
```

```
##      [,1] [,2] [,3]
## [1,]    0    0    0
## [2,]    2    0    0
## [3,]   14   14    0
```

Create parameter matrix for the PCC c13-c23-c12;3

```r
par13<-r13$par
par23<-r23$par
par12.3<-r12.3$par
par <- c(0, par12.3, par12,
            0, 0, par23,
            0, 0, 0)
par<- matrix(par, 3, 3)
par2<-matrix(0,3,3)
par2[2,1]<-r12.3$par2

RVM <- RVineMatrix(Matrix = Matrix, family = family,
                    par = par, par2 = par2,
                    names = c("shuck", "vis", "shell"))
```

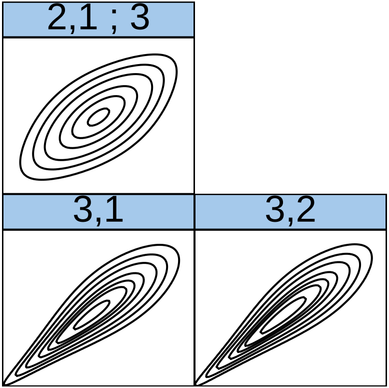
Fit R-vine with specified RVM c13-c23-c12;3

```
out12.3fixed<-RVineSeqEst(udata.f3, RVM=RVM, method = "mle")
summary(out12.3fixed)
```

```
## tree   edge | family  cop   par   par2 |  tau   utd   ltd
## --------------------------------------------------------
##    1    3,1 |     14   SG  2.93   0.00 | 0.66     -  0.73
##         3,2 |     14   SG  3.21   0.00 | 0.69     -  0.76
##    2  2,1;3 |      2    t  0.62  12.06 | 0.43  0.11  0.11
## ---
## type: C-vine    logLik: 2253.25    AIC: -4498.5    BIC: -4477.8
## ---
## 1 <-> shuck,   2 <-> vis,   3 <-> shell
```

Fitted normalized contour plots for specified RVM (c13-c23-c12;3)

```
contour(out12.3fixed)
```



7. Comparison of the three specified R-vines

Output function

```
vine.out<-function(fit=out12.3fixed,data=udata.f3,digits=2){
df<-sum(abs(fit$par)>0)+sum(fit$par2>0)
out<-round(c(RVineLogLik(data, fit)$loglik,df,
             RVineAIC(data,fit)$AIC,
             RVineBIC(data,fit)$BIC),digits)
names(out)<-c("loglik","par","AIC","BIC")
out
}
```

Apply output function

```
out.table<-rbind(
vine.out(fit=out13.2fixed,data=udata.f3),
vine.out(fit=out23.1fixed,data=udata.f3),
vine.out(fit=out12.3fixed,data=udata.f3))
row.names(out.table)<-c("c12-c23-c13;2",
                        "c13-c12-c23;1",
                        "c13-c23-c12;3")
out.table
```

```
##                loglik par      AIC      BIC
## c12-c23-c13;2 2253.91   3 -4501.82 -4486.29
## c13-c12-c23;1 2247.47   3 -4488.95 -4473.42
## c13-c23-c12;3 2253.25   4 -4498.50 -4477.80
```

8. Finding best R-vine using the step wise Dissmann algorithm

Select R-vine using the Dissmann algorithm

```
out3.select<-RVineStructureSelect(udata.f3, family=c(1,2,3,4), method = "mle")
summary(out3.select)
```

```
## tree   edge | family  cop   par  par2 |  tau   utd   ltd
## ------------------------------------------------------
##    1    2,1 |     14   SG  3.65  0.00 | 0.73     -  0.79
##         3,2 |     14   SG  3.21  0.00 | 0.69     -  0.76
##    2  3,1;2 |     14   SG  1.23  0.00 | 0.19     -  0.24
## ---
## type: C-vine    logLik: 2253.91    AIC: -4501.82    BIC: -4486.29
## ---
## 1 <-> shuck,   2 <-> vis,   3 <-> shell
```