Parametric and non parametric vine analysis of the Magic data (Application to Lecture 3)

Claudia Czado and Thomas Nagler, Technical University Munich

10 August, 2018

Setup

required packages and colors

```
library(VineCopula)
library(kdevine)
library(kdecopula)
library(rafalib)
library(dplyr)
```

```
##
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
##
##     filter, lag
```

```
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
cols <- naglr::naglr_pal()(5)
```

Variable description

Attribute Information:

1. Length: continuous # major axis of ellipse [mm]
2. Width: continuous # minor axis of ellipse [mm]
3. Size: continuous # 10-log of sum of content of all pixels [in #phot]
4. Conc: continuous # ratio of sum of two highest pixels over fSize [ratio]
5. Conc1: continuous # ratio of highest pixel over fSize [ratio]
6. Asym: continuous # distance from highest pixel to center, projected onto major axis [mm]
7. M3Long: continuous # 3rd root of third moment along major axis [mm]
8. M3Trans: continuous # 3rd root of third moment along minor axis [mm]
9. Alpha: continuous # angle of major axis with vector to origin [deg]
10. Dist: continuous # distance from origin to center of ellipse [mm]
11. class: g,h # gamma (signal), hadron (background)

g = gamma (signal): 12332 h = hadron (background): 6688

Read in data

```
dat <- readr::read_csv("magic04.data", col_names = FALSE)
```

```
## Parsed with column specification:
## cols(
##   X1 = col_double(),
##   X2 = col_double(),
##   X3 = col_double(),
##   X4 = col_double(),
##   X5 = col_double(),
##   X6 = col_double(),
##   X7 = col_double(),
##   X8 = col_double(),
##   X9 = col_double(),
##   X10 = col_double(),
##   X11 = col_character()
## )
```

```
names(dat) <- Hmisc::Cs(Length, Width, Size, Conc, Conc1, Asym,
    M3Long, M3Trans, Alpha, Dist, class)
dim(dat)
```

```
## [1] 19020    11
```

```
table(dat[,"class"])
```

```
##
##     g     h
## 12332  6688
```

Separate into observation with “g” and “h” class

```
udatg <- dat %>%
    dplyr::filter(class == "g") %>%
    dplyr::select(-class) %>%
    pobs()
dim(udatg)
```

```
## [1] 12332    10
```

```
udath <- dat %>%
    dplyr::filter(class == "h") %>%
    dplyr::select(-class) %>%
    pobs()
dim(udath)
```

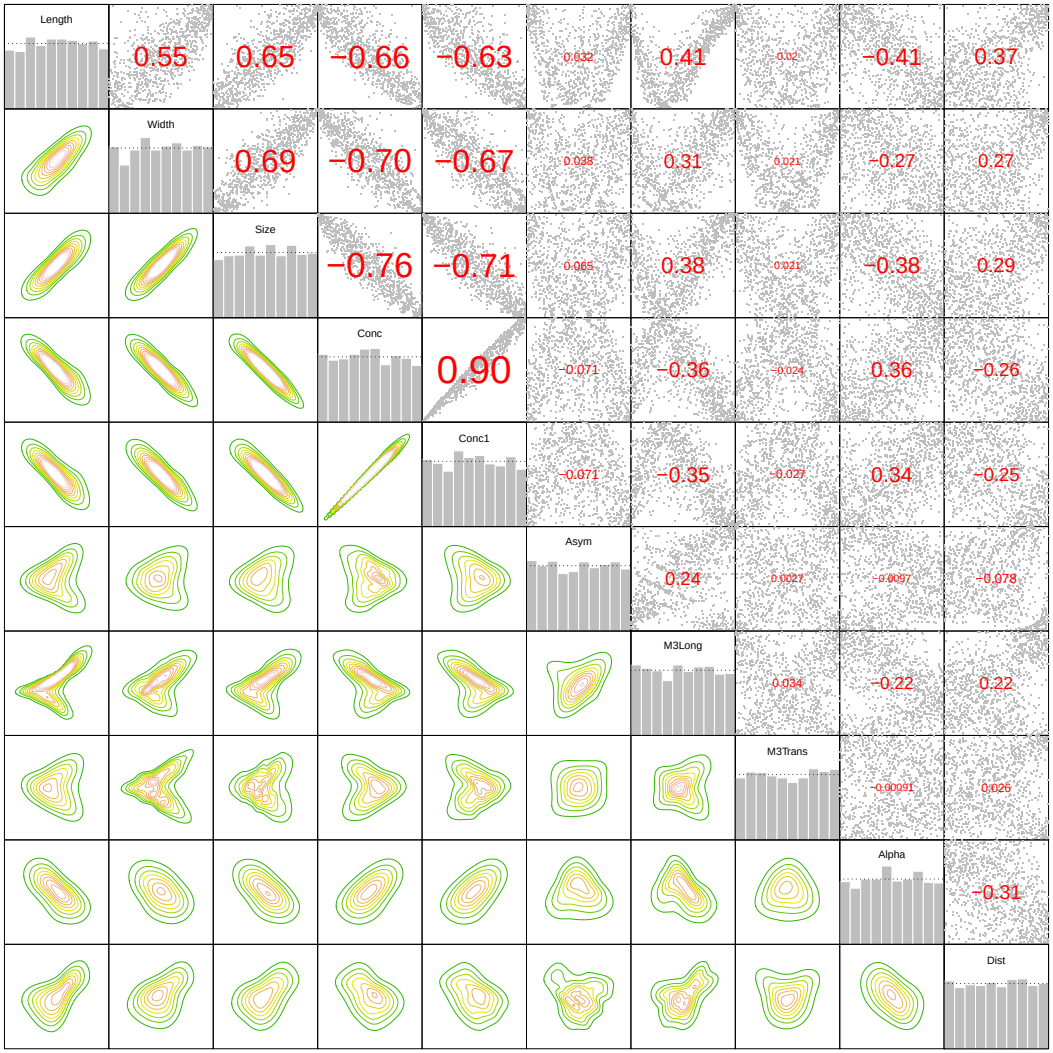
```
## [1] 6688   10
```

Take the first 1000 observations from both subsets

```
udatg1000<-as.copuladata(udatg[1:1000,])
udath1000<-as.copuladata(udath[1:1000,])
```

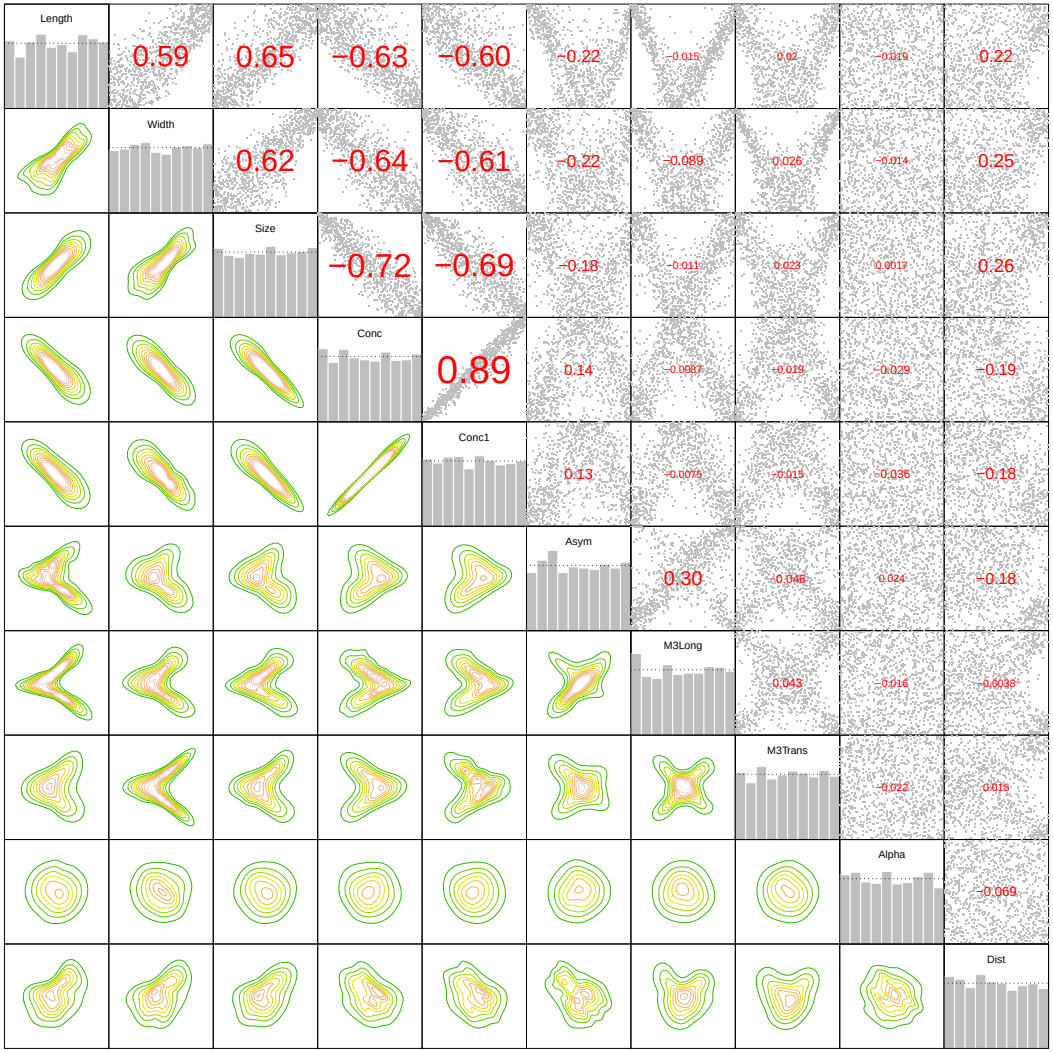
Normalized contour plots for class g

```
pairs(udatg1000)
```



Normalized contour plots for class h

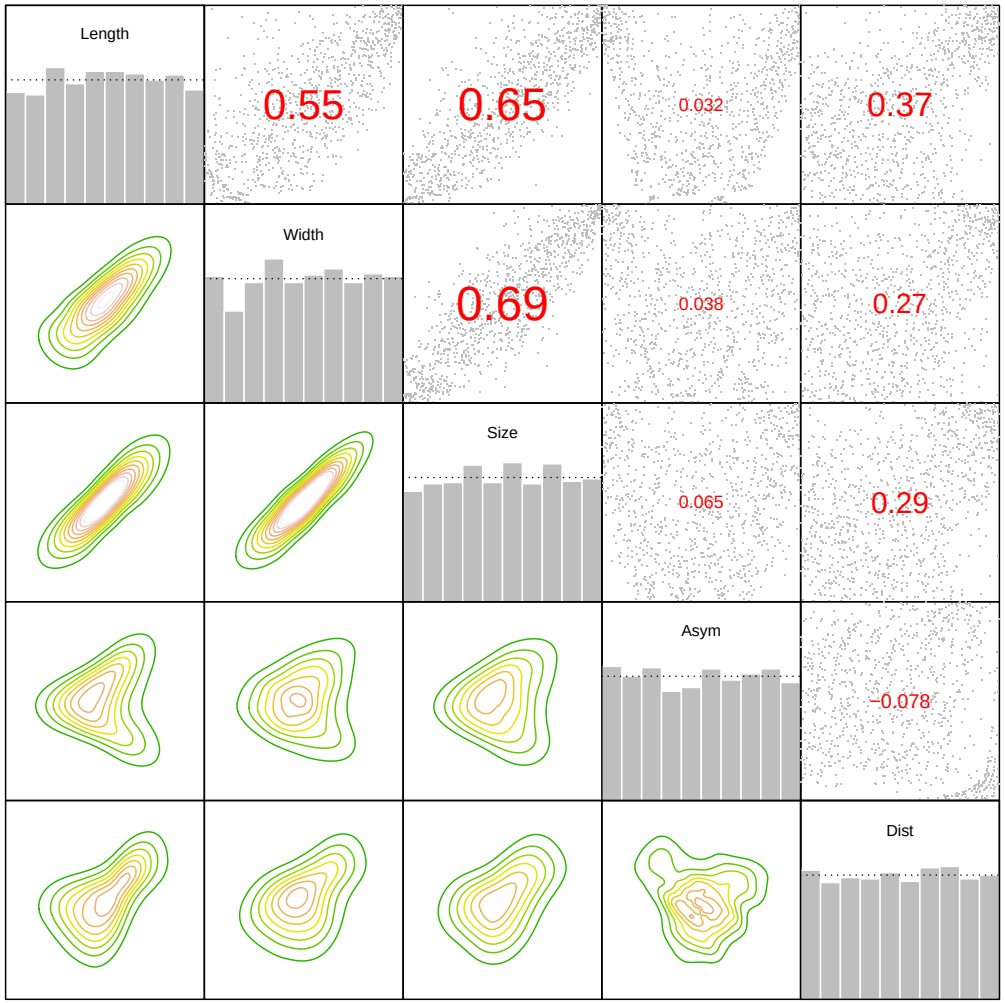
```
pairs(udath1000)
```



Analysis of the variables Length, Width, Size, Asym, Dist for class g

Pairs plot of the 5 variables for class g

```
udat5g<-udatg1000[,c("Length","Width","Size", "Asym", "Dist")]
pairs(udat5g)
```



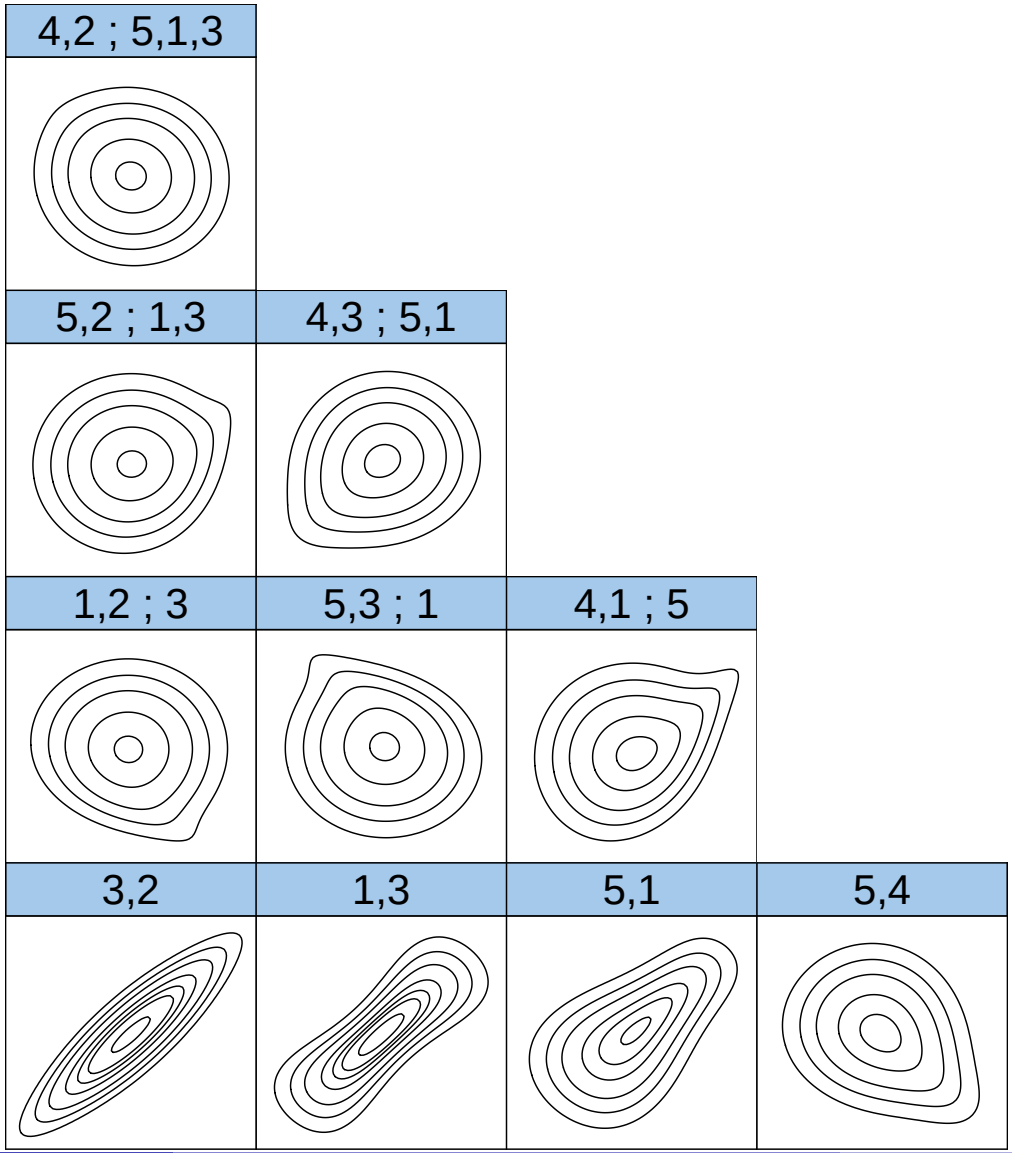
Fit parametric vine model for class g

```
rv.5g<-RVineStructureSelect(udat5g)
summary(rv.5g)
```

```
## tree       edge | family        cop    par    par2 |   tau   utd   ltd
## ------------------------------------------------------------------------
##    1        3,2 |      2          t   0.87   17.82 |  0.68  0.27  0.27
##             1,3 |      5          F   9.39    0.00 |  0.65     -     -
##             5,1 |     10        BB8   2.68    0.92 |  0.40     -     -
##             5,4 |     40    BB8_270  -1.39   -0.98 | -0.16     -     -
##    2       1,2;3 |    234  Tawn2_270  -1.45    0.09 | -0.06     -     -
##            5,3;1 |    224   Tawn2_90  -1.49    0.13 | -0.08     -     -
##            4,1;5 |    204      Tawn2   1.63    0.26 |  0.15  0.19     -
##    3     5,2;1,3 |    204      Tawn2   1.37    0.10 |  0.05  0.07     -
##          4,3;5,1 |     20       SBB8   1.23    0.97 |  0.09     -     -
##    4   4,2;5,1,3 |     24        G90  -1.03    0.00 | -0.03     -     -
## ---
## type: R-vine    logLik: 1643.87    AIC: -3251.74    BIC: -3163.4
## ---
## 1 <-> Length,   2 <-> Width,   3 <-> Size,   4 <-> Asym,   5 <-> Dist
```

Fitted contour plots of parametric model for class g

```
contour(rv.5g)
```



Function to extract model selection criteria from non parametric model fit

```
np.model<-function(fit=rv.5h.np,
          name="rv.5h.np"){data.frame(name=name,
          loglik=fit$info$loglik,effp=fit$info$effp,
          AIC=fit$info$AIC,BIC=fit$info$BIC)}
```

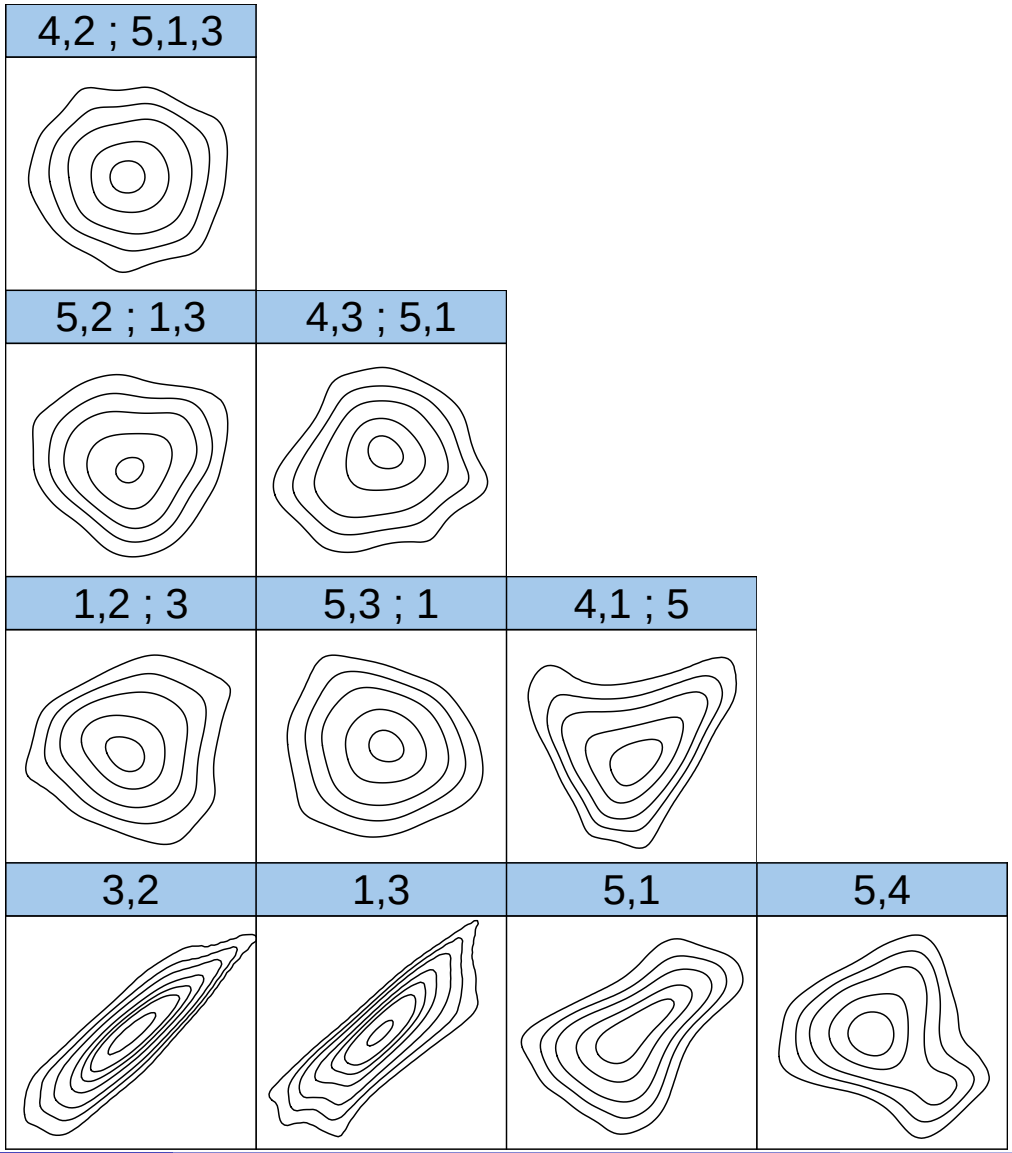
Fit nonparametric vine model for class g

```
rv.5g.np<-kdevinecop(udat5g,info=TRUE)
np.model(fit=rv.5g.np,name="rv.5g.np")
```

```
##       name   loglik     effp       AIC       BIC
## 1 rv.5g.np 2197.923 215.0544 -3965.736 -2910.302
```

Fitted contours of nonparametric vine modelfor class g

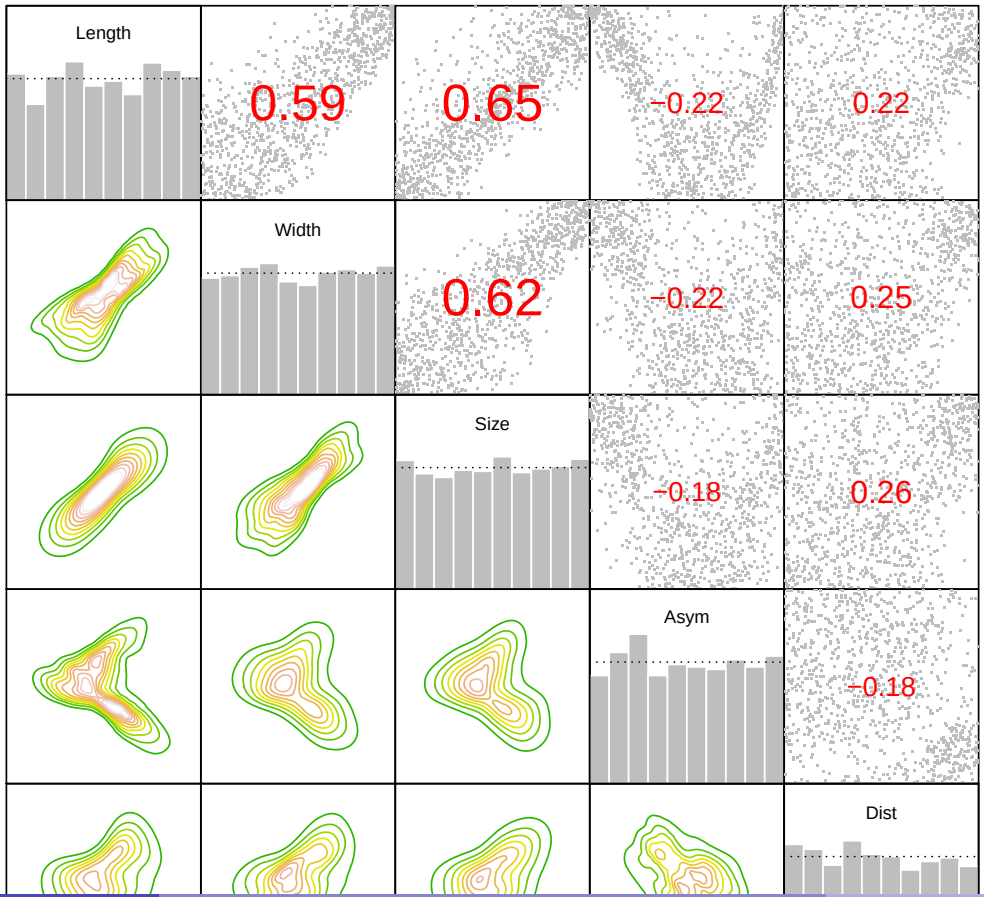
```
contour(rv.5g.np)
```



Analysis of the variables Length, Width, Size, Asym, Dist for class h

Contour plots of the variables Length, Width, Size, Asym, Dist for class h

```
udat5h<-udath1000[,c("Length","Width","Size", "Asym", "Dist")]
pairs(udat5h)
```



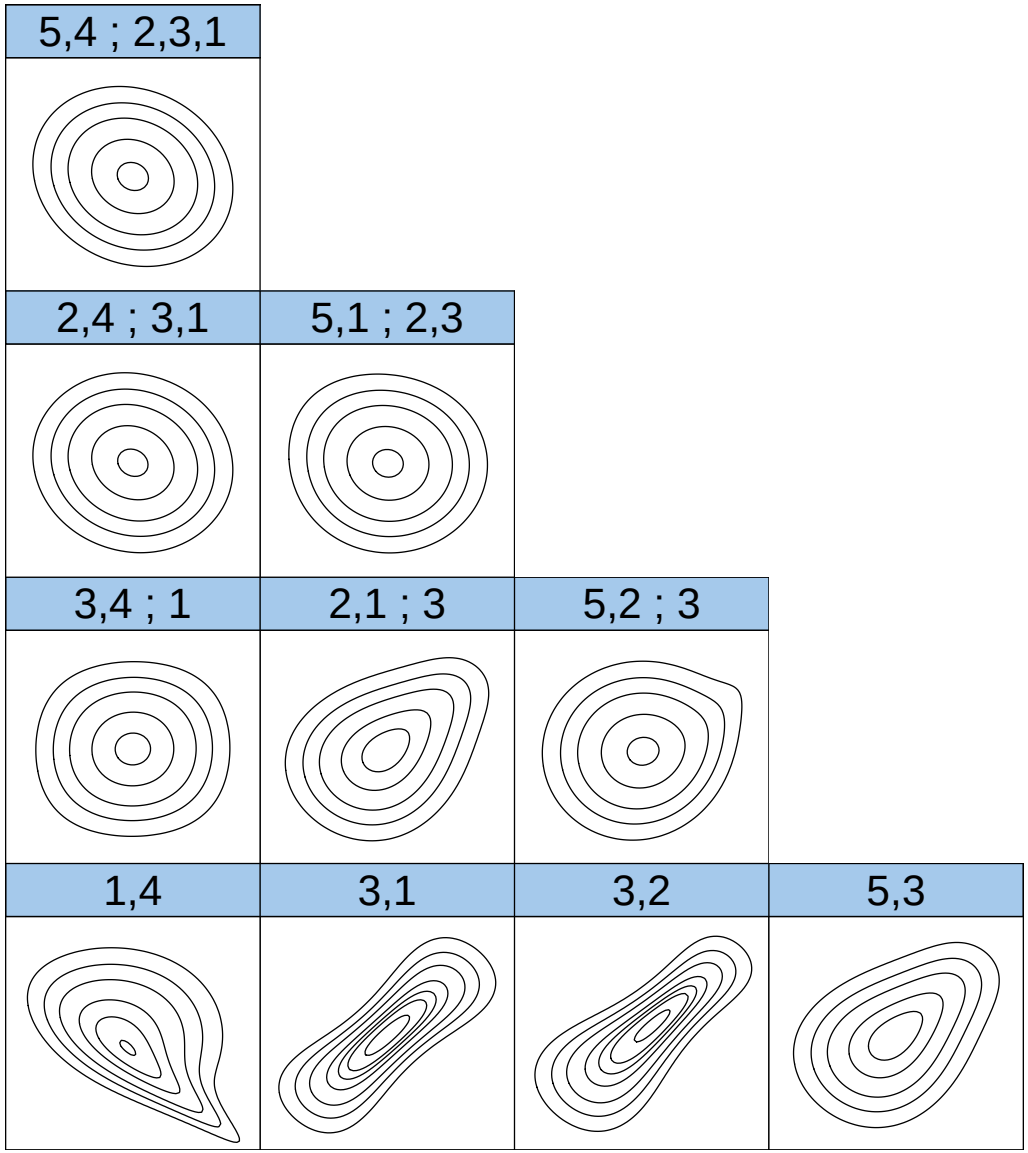
Fit parametric vine model for class h

```
rv.5h<-RVineStructureSelect(udat5h)
summary(rv.5h)
```

```
## tree       edge | family        cop    par   par2 |   tau   utd   ltd
## ------------------------------------------------------------------------
##    1        1,4 |    234  Tawn2_270  -2.30   0.42 | -0.30     -     -
##             3,1 |      5          F   9.23   0.00 |  0.64     -     -
##             3,2 |     10        BB8   6.00   0.81 |  0.61     -     -
##             5,3 |     10        BB8   2.05   0.90 |  0.27     -     -
##    2      3,4;1 |      2          t   0.01  14.53 |  0.01  0.00  0.00
##           2,1;3 |     10        BB8   1.71   0.94 |  0.22     -     -
##           5,2;3 |    204      Tawn2   1.31   0.15 |  0.07  0.09     -
##    3    2,4;3,1 |      5          F  -0.49   0.00 | -0.05     -     -
##         5,1;2,3 |     33       C270  -0.07   0.00 | -0.04     -     -
##    4  5,4;2,3,1 |      1          N  -0.13   0.00 | -0.09     -     -
## ---
## type: R-vine    logLik: 1534.81    AIC: -3037.62    BIC: -2959.1
## ---
## 1 <-> Length,   2 <-> Width,   3 <-> Size,   4 <-> Asym,   5 <-> Dist
```

Fitted contour plots of parametric model for class h

```
contour(rv.5h)
```



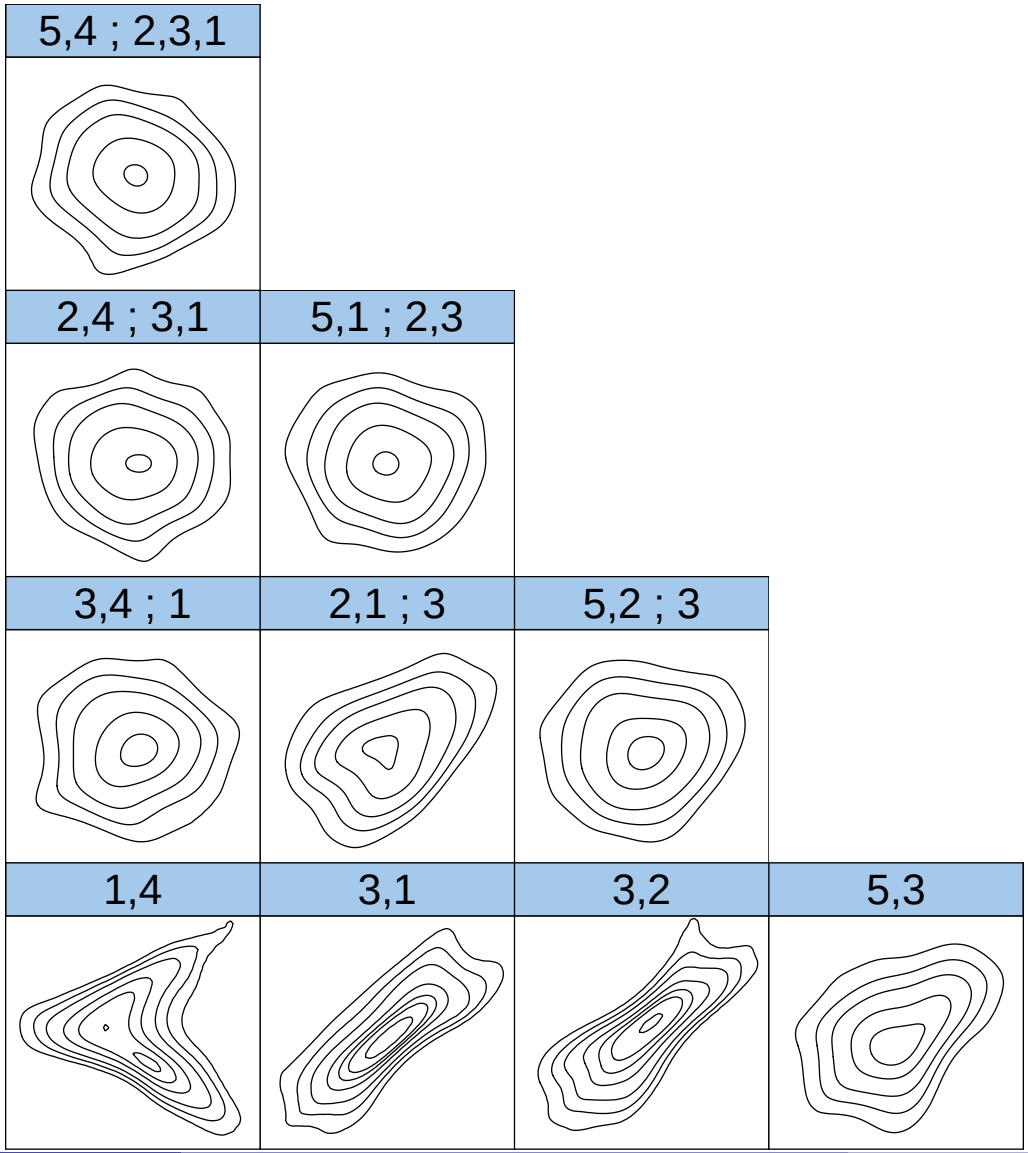
Fit nonparametric vine model for class h

```
rv.5h.np<-kdevinecop(udat5h,info=TRUE)
np.model(rv.5h.np)
```

```
##        name   loglik     effp       AIC       BIC
## 1 rv.5h.np 2128.055 206.7949 -3842.521 -2827.622
```

Fitted contour of nonparametric vine model for class h

```
contour(rv.5h.np)
```



Truncated models

Parametric with independence tests for class g

```
rv.5g.ind<-RVineStructureSelect(udat5g,indeptest = TRUE)
summary(rv.5g.ind)
```

```
## tree       edge | family       cop    par   par2 |   tau   utd   ltd
## ------------------------------------------------------------------
##    1        3,2 |      2         t   0.87  17.82 |  0.68  0.27  0.27
##             1,3 |      5         F   9.39   0.00 |  0.65     -     -
##             5,1 |     10       BB8   2.68   0.92 |  0.40     -     -
##             5,4 |     40   BB8_270  -1.39  -0.98 | -0.16     -     -
##    2      1,2;3 |      0         I      -      - |  0.00     -     -
##           5,3;1 |    224  Tawn2_90  -1.49   0.13 | -0.08     -     -
##           4,1;5 |    204     Tawn2   1.63   0.26 |  0.15  0.19     -
##    3    5,2;1,3 |      0         I      -      - |  0.00     -     -
##         4,3;5,1 |     20      SBB8   1.23   0.97 |  0.09     -     -
##    4  4,2;5,1,3 |      0         I      -      - |  0.00     -     -
## ---
## type: R-vine    logLik: 1624.81    AIC: -3223.62    BIC: -3159.82
## ---
## 1 <-> Length,   2 <-> Width,   3 <-> Size,   4 <-> Asym,   5 <-> Dist
```

Parametric with independence tests for class h

```
rv.5h.ind<-RVineStructureSelect(udat5h,indeptest = TRUE)
summary(rv.5h.ind)
```

```
## tree       edge | family        cop    par  par2 |   tau   utd   ltd
## ------------------------------------------------------------------
##    1        1,4 |    234  Tawn2_270  -2.30  0.42 | -0.30     -     -
##             3,1 |      5          F   9.23  0.00 |  0.64     -     -
##             3,2 |     10        BB8   6.00  0.81 |  0.61     -     -
##             5,3 |     10        BB8   2.05  0.90 |  0.27     -     -
##    2      3,4;1 |      0          I      -     - |  0.00     -     -
##           2,1;3 |     10        BB8   1.71  0.94 |  0.22     -     -
##           5,2;3 |    204      Tawn2   1.31  0.15 |  0.07  0.09     -
##    3    2,4;3,1 |      5          F  -0.48  0.00 | -0.05     -     -
##         5,1;2,3 |      0          I      -     - |  0.00     -     -
##    4  5,4;2,3,1 |    224   Tawn2_90  -1.30  0.12 | -0.06     -     -
## ---
## type: R-vine    logLik: 1533.37    AIC: -3038.73    BIC: -2970.02
## ---
## 1 <-> Length,   2 <-> Width,   3 <-> Size,   4 <-> Asym,   5 <-> Dist
```

Non parametric with independence tests for class g

```
rv.5g.np.ind<-kdevinecop(udat5g,test.level=.1,info=TRUE)
np.model(rv.5g.np.ind)
```

```
##       name   loglik     effp       AIC       BIC
## 1 rv.5h.np 2088.903 152.7898 -3872.227 -3122.372
```

Non parametric with independence tests for class h

```
rv.5h.np.ind<-kdevinecop(udat5h,test.level=.1,info=TRUE)
np.model(rv.5h.np.ind)
```

```
##       name   loglik     effp       AIC       BIC
## 1 rv.5h.np 2064.749 145.6357 -3838.226 -3123.482
```