

Copula-based directional dependence networks for multivariate data

Subho Majumdar

Stat 8931 course project
School of Statistics, University of Minnesota- Twin Cities
e-mail: majum010@umn.edu

Abstract. Maximum likelihood-depth estimates for Clayton and t -copula parameters are derived and unbiased estimates for true parameters are derived from them using simulation. After that an algorithm is formulated which selects best-fitting copula from a choice of Gaussian, t , Clayton and Gumbel copula families by minimizing euclidean distance with the empirical copula for each variable-pair of a multivariate data matrix. Maximum likelihood and maximum likelihood depth are used for obtaining copula parameter estimates. Two real data applications are also given.

Keywords : Copula; Directional dependence; Likelihood depth

1 Introduction

Copula functions are used to model joint dependency structure in financial and survival data, but their applications have been limited for multivariate data mostly due to limitation of parametric models. A tree structure called *Vine* made up of conditional bivariate copulae can be used to model multivariate dependence problems, but it does not give a unique decomposition of the joint multivariate density [2]. An important part of fitting copula models to a bivariate data is to first choose the copula that gives the best fit. Some empirical methods of doing this are using AIC and BIC [14] as well as minimizing distance to the empirical copula function [13].

Estimates of copula functions can be obtained using parametric, semiparametric [6][7] and non-parametric (empirical copula [3]) methods. Apart from using maximum likelihood estimates, an alternative form of parametric estimation is by maximizing the likelihood depth function for a known copula family. Likelihood-based depth functions [11] are a consequence of the generalized theory of depth put forward by Mizera [10]. For bivariate Gaussian and Gumbel copulae, likelihood depth-based unbiased estimates of the corresponding parameter have been derived and tests based on them have been formulated [5].

Although estimation of pairwise copulae in multivariate genetic data has been explored before, it did not involve copula selection [8]. Here we start with some general notions related to copula functions and likelihood depth in Section 2. In Section 3 we first obtain unbiased maximum likelihood-depth estimates (MLDE) for Clayton and t -copula families. After that for data given on two variables, copula parameter estimates are obtained for four choices of copulae, either by ML or MLD method. In each case best-fitting copula is found by minimizing the euclidean distance to the empirical copula. Repeating this process for all variable-pairs in a multivariate data matrix yields a network that characterizes tail dependencies between each pair of variables, due to our choice of the copula families considered. Finally in Section 4 the method is applied to two datasets to get an idea of its outcomes.

2 Preliminaries

2.1 Copulas

A copula is any function $C : [0, 1]^2 \rightarrow [0, 1]$ has uniformly distributed marginals, and according to Sklar's theorem, for any bivariate distribution function H on random variables X, Y having marginals F, G respectively, there always exists a copula function C such that

$$H(x, y) = C(F(x), G(y))$$

In other words we have $C(u, v) = H(F^{-1}(u), G^{-1}(v))$ for $(u, v) \in [0, 1]^2$.

Copula functions characterize the dependence structure between two (or more) random variables, regardless of their marginal distributions. Different kinds of joint tail behaviors are captured using different copula families [1]:

Gaussian Copula This arises from a bivariate normal distribution with standard normal marginals, and has distribution function

$$C_\rho(u, v) = \int_{-\infty}^{\Phi^{-1}(u)} \int_{-\infty}^{\Phi^{-1}(v)} \frac{1}{2\pi\sqrt{1-\rho^2}} \exp\left\{-\frac{x^2 - 2xy + y^2}{2(1-\rho^2)}\right\} dx dy$$

The only parameter is $\rho \in [-1, 1]$, the correlation coefficient. Gaussian copula is characteristic of bivariate distributions with little probability of joint extreme events.

t-copula Used to model joint heavy-tail behavior, the *t*-copula also has the correlation coefficient as its parameter, but additionally depends on the degree of freedom ν as well. The copula is given by

$$C_{\rho, \nu}(u, v) = \int_{-\infty}^{t_\nu^{-1}(u)} \int_{-\infty}^{t_\nu^{-1}(v)} \frac{1}{2\pi\sqrt{1-\rho^2}} \left\{-\frac{x^2 - 2xy + y^2}{2(1-\rho^2)}\right\}^{-(\nu+2)/2} dx dy$$

A *t*-copula with $\nu \geq 30$ is considered equivalent to Gaussian copula.

Clayton copula Unlike the above two copulas which do not have a closed form expressions (implicit copulas), there are explicit copulas with simple forms for the distribution functions. Clayton copula is one such explicit copula, and is relevant when the random variables have more dependence for joint smaller values but little dependence for joint large values. The copula is given by

$$C_\delta(u, v) = (u^{-\delta} + v^{-\delta} + 1)^{-1/\delta}$$

Here $\delta \in (0, \infty)$ is the dependency parameter, with $\delta \rightarrow 0$ and $\delta \rightarrow \infty$ implying complete independence and dependence, respectively.

Gumbel copula This explicit copula exhibits strong positive tail dependence but little negative tail dependence.

$$C_\delta(u, v) = \exp\left[-\left\{(-\log u)^\delta + (-\log v)^\delta\right\}^{1/\delta}\right]$$

The copula parameter is $\delta \in [1, \infty)$. $\delta = 1$ implies independence, while $\delta \rightarrow \infty$ means perfect dependence.

Note: Both Clayton and Gumbel copulae are used when the random variables are positively dependent on each other. For negative dependence, their 90° (or 270°) rotated versions are used, which have the same expressions for $C_\delta(u, v)$, but $\delta \in (-\infty, 0)$ for Clayton and $\in (-\infty, 1]$ for Gumbel copula, respectively.

2.2 Likelihood depth

Consider n iid observations from $Z \sim f_\theta$ with $\theta \in \Theta \subset \mathbb{R}^p$. Then, we have the likelihood function $L(\theta, z) = f(\theta, z)$ and we can define the following [10][11]

Nonfit A parameter $\theta \in \Theta$ is called nonfit with respect to data (z_1, \dots, z_n) when there exists $\theta' \neq \theta \in \Theta$ such that $L(\theta', z_i) > L(\theta, z_i)$, or equivalently $l_i(\theta') > l_i(\theta)$ for $i = 1, \dots, n$. (Here $l_i(\theta) = \log(L(\theta, x))$ for any $\theta \in \Theta$)

Likelihood depth Likelihood depth at a parameter value θ is the minimum proportion of observations that need to be deleted from the data to make θ a nonfit.

Tangent Likelihood depth Denoted by $d_T(\theta, \mathbf{z})$, it is defined as

$$d_T(\theta, \mathbf{z}) = \frac{1}{n} \inf_{u \neq \mathbf{0}_p} \#\{i : u^T \nabla l_i(\theta) \leq 0\}$$

Evidently $d_T(\theta, \mathbf{z}) \leq 1/2$. When the parameter space is 1-dimensional, i.e. $p = 1$, the tangent likelihood depth simplifies to

$$d_T(\theta, \mathbf{z}) = \frac{1}{n} \left[\min\{N_+^\theta, N_-^\theta\} + N_0^\theta \right]$$

with $N_+^\theta = \#\{i : l'_i(\theta) > 0\}$, $N_-^\theta = \#\{i : l'_i(\theta) < 0\}$ and $N_0^\theta = \#\{i : l'_i(\theta) = 0\}$.

The tangent likelihood depth is equal to likelihood depth under very general conditions [10]. For this reason, from now on by likelihood depth we shall mean tangent likelihood depth.

Maximum Likelihood Depth estimator (MLDE) It is a parameter value (not necessarily unique) that maximizes the likelihood depth over the parameter space Θ :

$$\hat{\theta}_d \in \arg \max_{\theta \in \Theta} d_T(\theta, \mathbf{z})$$

Under certain regularity conditions, depth at each $\theta \in \Theta$ i.e. $d_T(\theta, \mathbf{z})$ converges uniformly to its population analogue $d_T(\theta, P)$, where P is a valid probability measure [10][11]. Given the point with maximum depth is unique in the population, say θ_d , we also have $\hat{\theta}_d \rightarrow \theta_d$ uniformly [4].

3 Methods

3.1 Estimation of copula parameters

For all the copulas described above, there are 1-1 mappings between the copula parameter and Kendall's τ [1]. Thus one way to obtain parameter estimates is inverting from the sample version of Kendall's τ . One can also obtain maximum likelihood estimates of the copula parameters. For Gaussian copula there is a closed form solution of the sample correlation matrix, while for the other 3, MLEs are obtained by iteratively maximizing the log-likelihood. For t -copula the degree of freedom is estimated by numerical search after obtaining the estimate of correlation parameter [9].

Maximum likelihood depth points obtained from copula log-likelihoods give biased estimates of the true parameter for Gaussian and Gumbel copula [5]. The inverse mapping from the biased estimate to the actual parameter value is obtained by simulating over a range of values. This relation turns out to be cubic for Gaussian copula:

$$\rho = \begin{cases} \text{sign}(\hat{\rho}_d) [-1.22217|\hat{\rho}_d|^3 + 3.6434|\hat{\rho}_d|^2 - 1.42154|\hat{\rho}_d| + 0.000396] & \text{if } |\hat{\rho}_d| > 0.461 \\ 0 & \text{if } |\hat{\rho}_d| \leq 0.461 \end{cases}$$

and linear for Gumbel copula: $\delta = 0.02 + 0.706\hat{\delta}_d$.

We use the same simulation setup as in [5] (1000 samples of size 100 each) and interval search algorithm (starting interval $[0, 0.99]$ for t -copula, $[0.1, 10\hat{\delta}_{ML}]$ for Clayton) to obtain MLDEs for a range of parameter values for Clayton and t -copula (df 5 and 10). For all the cases the average of maximum depth points are different from the true parameters. After this we perform a larger simulation to numerically obtain the relation between biased MLDEs and true parameters. For Clayton copula it is linear:

$$\delta = -0.5302 + 0.7163\hat{\delta}_d$$

while for t -copulae the relation is a cubic equation, with different coefficients depending on the df-parameter ν (listed in appendix, $\nu = 3, 4, \dots, 30$). Thus depth-based unbiased estimators of parameters for all the copulae considered here can be obtained from the respective MLDEs.

3.2 Copula selection

Given data $\mathbf{Z}_1, \dots, \mathbf{Z}_n$ with $\mathbf{Z}_i = (X_i, Y_i)$ we first calculate the empirical copula of the data:

$$C_e(u, v) = \frac{1}{n} \sum_{i=1}^n I(U_i \leq u, V_i \leq v)$$

where $\mathbf{U} = F_n(\mathbf{X})$, $\mathbf{V} = G_n(\mathbf{Y})$ are pseudosamples obtained from the data using marginal empirical distributions F_n, G_n respectively. Given any copula C , we can calculate its distance from the empirical copula:

$$d(C, C_e) = \sqrt{\sum_{i=1}^n \sum_{j=1}^n \left[C\left(\frac{i}{n}, \frac{j}{n}\right) - C_e\left(\frac{i}{n}, \frac{j}{n}\right) \right]^2}$$

For data on a pair of variables, we first calculate Kendall's τ , and if it is > 0 (or < 0) consider Gaussian, t , (rotated) Clayton and (rotated) Gumbel copulae. ML and depth-based estimates are obtained for all copula parameters and then plugged into corresponding distributions. Finally for each method, ML and depth-based, the copula with least distance from the empirical copula as calculated above is chosen as the best-fitting copula. The degree of freedom parameter for both the methods is set as the one estimated by ML method.

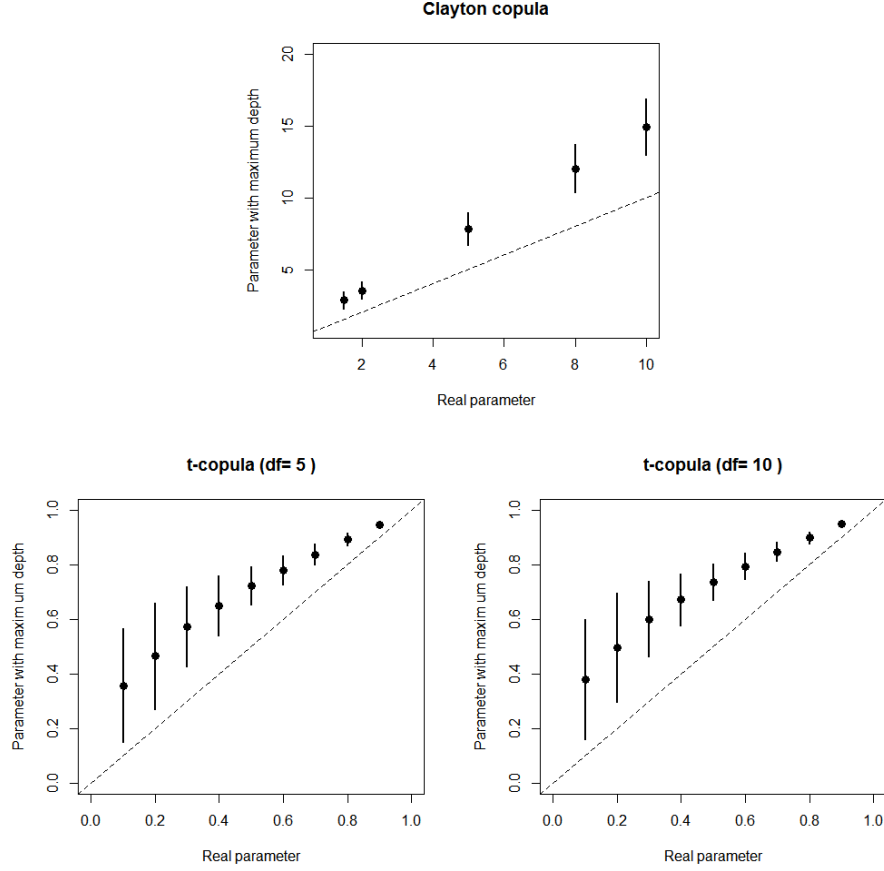


Fig. 1. Mean and standard deviations of simulated parameters with max depth

3.3 Constructing the graph

The Kendall's τ calculated from a bivariate dataset $(\mathbf{Z}_1, \dots, \mathbf{Z}_n)$ with $\mathbf{Z}_i = (X_i, Y_i); i = 1, \dots, n$ has an asymptotic normal distribution under the null hypothesis that X and Y are independent:

$$\hat{\tau} \sim AN \left(0, \frac{2(2n+5)}{9n(n-1)} \right)$$

For a given multivariate data matrix, we consider a pair of variables, and given the sample size n is large enough, test for independence between the variables using the asymptotic test above. Given that there is evidence of dependence, we attempt to choose the copula function that is the best fit to the data using the method outlined in 3.2, thus ending up with two copulas functions for ML and depth-based parameter estimations. Finally the whole process is repeated for all pairs of variables in the data.

Algorithm 1 Algorithm to obtain pairwise copula dependence network

```
1: procedure COPNETWORK(data matrix  $\mathbf{D} \in \mathbb{R}^{n \times p}$ )
2:   Set  $i = 1, j = 1$ .
3:   top:
4:   Set  $\mathbf{X} = i^{th}$  column of  $\mathbf{D}$ ,  $\mathbf{Y} = j^{th}$  column of  $\mathbf{D}$ .
5:   Check for independence of  $\mathbf{X}$  and  $\mathbf{Y}$  using test above.
6:   if Independent then
7:     goto update
8:   else
9:     if Kendall's  $\tau > 0$  then
10:       $S = \{\text{Gaussian}, t, \text{Clayton}, \text{Gumbel}\}$ 
11:     else
12:       $S = \{\text{Gaussian}, t, \text{rotated Clayton}, \text{rotated Gumbel}\}$ 
13:      Select  $C_{ij,M} \in S$  as the best fitting copula, parameter estimated by ML method.
14:      Select  $C_{ij,D} \in S$  as the best fitting copula, parameter estimated by depth-based method.
15:   update:
16:   if  $j = p$  then
17:     if  $i = p$  then Stop
18:     else
19:       Set  $i \leftarrow i + 1, j \leftarrow i$ , goto top
20:   else
21:     Set  $j \leftarrow j + 1$ , goto top
```

At the end of the algorithm we end up with two graphs, \mathbf{C}_M and \mathbf{C}_D , giving best-fitting copulae, obtained by the two respective methods, for each pair of variables.

4 Results

All analyses were performed in R version 3.0.2 [12]. We use the R packages `VineCopula` and `copula` for copula-related calculations. After obtaining two matrices $\mathbf{C}_M, \mathbf{C}_D$ we visualize the directional dependence structure by plotting it as a graph. Nodes correspond to variables and edges between two nodes are colored according to the type of best-fitting copula between the variables.

We now apply the methodology outlined above in two datasets from the UCI Machine Learning Repository (<http://archive.ics.uci.edu/ml>) to study the nature of dependence between variables.

4.1 Breast Tissue data

The data consists of impedance measurements from 106 samples of freshly excised breast tissue. There are 9 measurement variables (I0, PA500, HFS, DA, Area, A/DA, Max IP, DR, P) and a class variable specifying the class of Breast Cancer the patient has (6 or 4 classes). For our analysis we ignore the class variable due to small sample sizes in each class and obtain the networks from the measurement variables only.

There are 28 significantly dependent variable pairs among 36 possible ones. Graphs indicate that variables except the two phase angle related variables: HFS and PA500, are all dependent on one another. Most of these dependencies are symmetric, and heavy-tailed (t -copula) as per the ML method, but light-tailed as per the depth-based method (Figure 2).

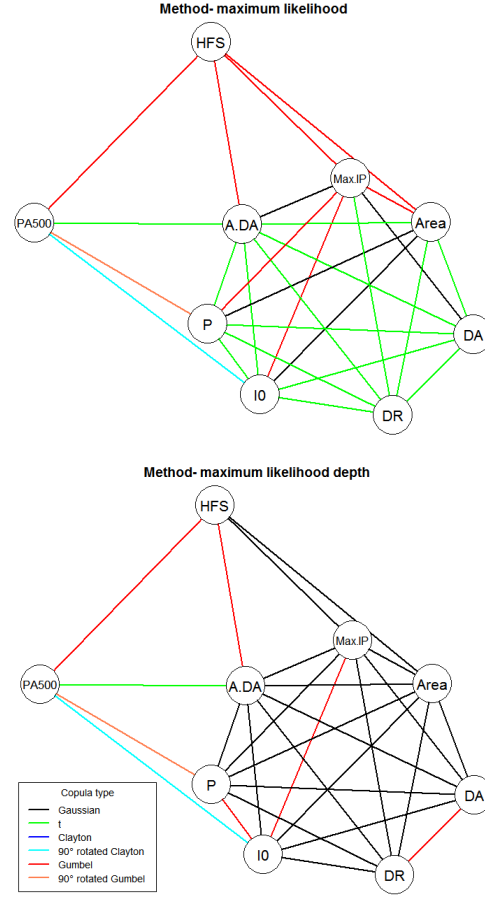


Fig. 2. Graph of Breast Cancer variables obtained by best-fitting (top) ML copula, (Bottom) MLD copula

| Variable name | Description |
|---------------|--|
| I0 | Impedivity (ohm) at zero frequency |
| PA500 | Phase angle at 500 KHz |
| HFS | High-frequency slope of phase angle |
| DA | Impedance distance between spectral ends |
| AREA | Area under spectrum |
| A/DA | Area normalized by DA |
| MAX IP | Maximum of the spectrum |
| DR | Distance between I0 and real part of the maximum frequency point |
| P | Length of the spectral curve |

Table 1. Description of measurement variables in Breast Cancer data

4.2 Cardiotocography data

This dataset has measurements relating to 2126 fetal cardiotocograms. There are 21 predictor variables, and a class variable describing fetal state (N=normal, S=suspect, P=pathologic). The descriptions of variables are given in the appendix. There are 3 types of variables:

1. Variables 1-7 give numerical measurements, like number of fetal movements, uterine contractions etc (Group A),
2. Variables 8-11 are about variability of cardiograms with respect to time (Group B),
3. Variables 12-21 give measurements of heart-rate histograms, like max, min, mean, median etc. (Group C)

DS, the sixth variable has most of its values set at 0, so we exclude it from our analysis.

Variable group-specific analysis Over 90% of connections between the 20 variables (176 among 190 possible) are found significant in the initial screening for dependence. Hence instead of plotting the graph we analyze the dependence structures within and between the 3 variable groups. There are 3 within-group (AA, BB, CC) and 3 between-group (AB, AC, BC) interactions. The type of copulas fit by ML method for each of these 6 interactions are listed in Table 2.

| Interaction | Independent | Gaussian | t | Clayton | rot. Clayton | Gumbel | rot. Gumbel |
|-------------|-------------|----------|-----|---------|--------------|--------|-------------|
| AA | 2 | 0 | 0 | 7 | 1 | 0 | 5 |
| BB | 1 | 2 | 1 | 0 | 1 | 0 | 1 |
| CC | 2 | 11 | 9 | 2 | 7 | 5 | 9 |
| AB | 1 | 4 | 3 | 2 | 11 | 1 | 2 |
| AC | 3 | 16 | 2 | 16 | 18 | 2 | 3 |
| BC | 5 | 7 | 3 | 7 | 11 | 2 | 5 |

Table 2. Summary of best-fitting copulae for within (Top 3) and between (bottom 3) group dependencies in CTG data

Dependence structure for different fetal classes We also compare between the within-group dependencies for the 3 fetal classes. Figs 3 and 4 outline differences in groups A and B for 3 sample classes. For variable group C, the summary of copulae fit between the 45 variable-pairs is summarized in table 3. Highlights include a high amount of independence in suspect class and high asymmetric dependencies in pathologic class.

| Sample class | Independent | Gaussian | t | Clayton | rot. Clayton | Gumbel | rot. Gumbel |
|--------------|-------------|----------|-----|---------|--------------|--------|-------------|
| Normal | 4 | 13 | 8 | 4 | 5 | 7 | 4 |
| Suspect | 11 | 8 | 8 | 3 | 3 | 10 | 2 |
| Pathologic | 3 | 9 | 2 | 2 | 11 | 7 | 11 |

Table 3. Best-fitting copulae by ML method for group C variables in CTG data

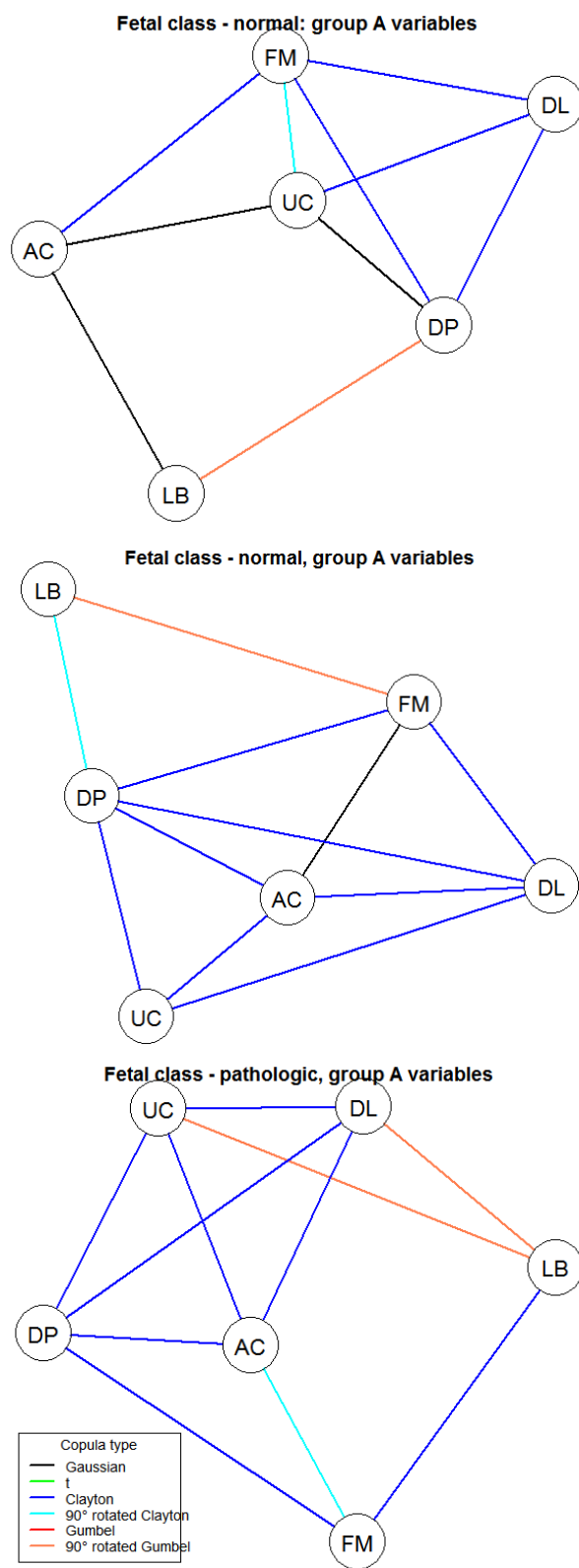


Fig. 3. Graph of group A variables for 3 classes

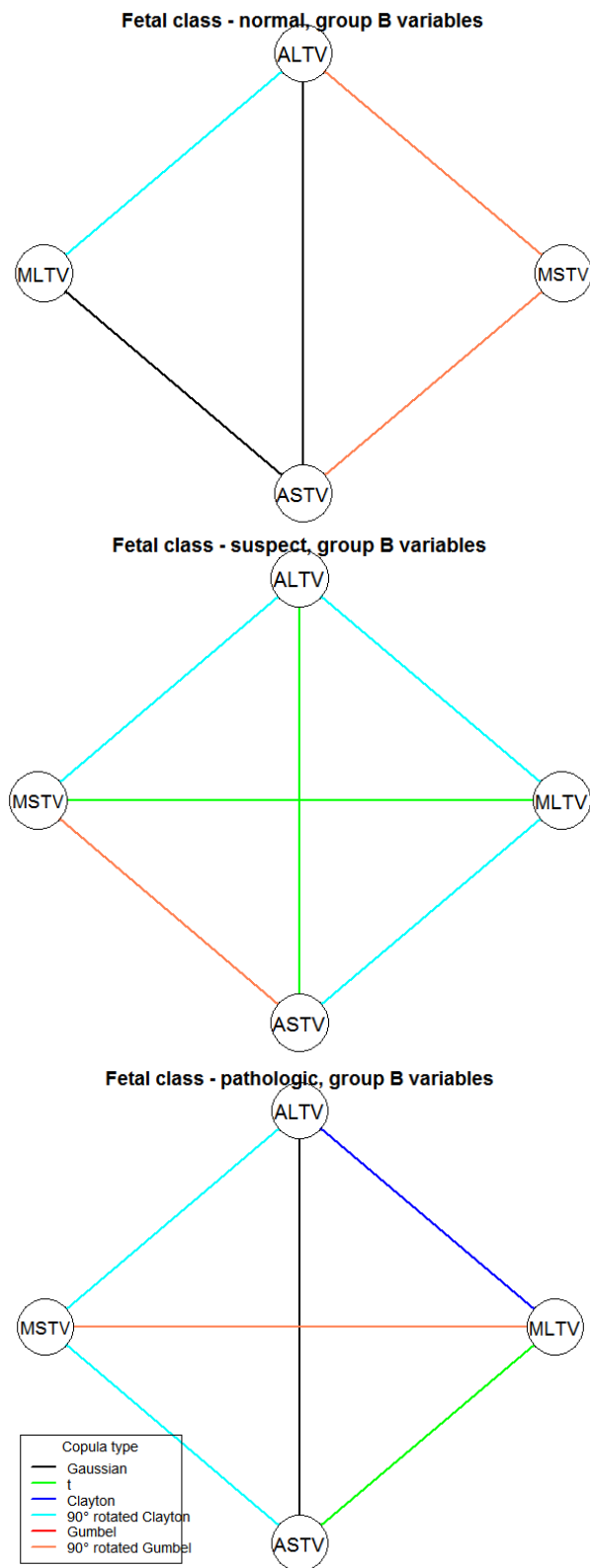


Fig. 4. Graph of group B variables for 3 classes

5 Conclusion

In the above sections we have outlined a distribution-free method to analyze the nature of dependencies between pairs of variables in a multivariate dataset. Applications on two real datasets are also done.

We aim to formulate a robust framework for estimating directional dependencies that can extend the applicability domain of copula functions. Further work on this direction includes using conditional bivariate copula for each variable-pair to eliminate effect of other variables. We also plan to do detailed simulation studies to compare between the two methods of copula parameter estimation, as well as analyze genetic data and compare the methodology with other known methods.

Acknowledgment

I thank my adviser Prof. Snigdhasu Chatterjee for his guidance and valuable inputs throughout the project.

References

1. AAS, K. Modelling the dependence structure of financial assets: A survey of four copulas. Tech. Rep. SAMBA/22/04, Norwegian Computing Center, Oslo, Norway, December 2004.
2. BRECHMANN, E., AND SCHEPMEIER, U. Modeling dependence with C- and D-vine copulas: The R package CDVine. *Journal of Statistical Software* 52 (2013), 1–27.
3. CAPÁRAÀ, P., FOUGÈRES, A.-L., AND GENEST, C. A nonparametric estimation procedure for bivariate extreme value copulas. *Biometrika* 84 (1997), 567–577.
4. DENECKE, L. *Estimators and Tests based on Likelihood-Depth with Application to Weibull Distribution, Gaussian and Gumbel Copula*. PhD thesis, University of Cassel, March 2010.
5. DENECKE, L., AND MÜLLER, C. Robust estimators and tests for bivariate copulas based on likelihood depth. *Computational Statistics and Data Analysis* 55 (2011), 2724–2738.
6. GENEST, C., GHOUDI, K., AND RIVEST, L.-P. A semiparametric estimation procedure in multivariate families of distribution. *Biometrika* 82 (1995), 543–552.
7. HOFF, P. Extending the rank likelihood for semiparametric copula estimation. *The Annals of Applied Statistics* 1 (2007), 265–283.
8. KIM, J.-M., JUNG, Y.-S., SUNGUR, E., HAAN, K.-H., PARK, C., AND SOHN, I. A copula method for modeling directional dependence of genes. *BMC Bioinformatics* 9 (2008).
9. MASHAL, R., AND ZEEVI, A. Beyond correlation: Extreme co-movements between financial assets. Tech. rep., Columbia University, 2002.
10. MIZERA, I. On depth and deep points: A calculus. *The Annals of Statistics* 30 (2002), 1681–1736.
11. MIZERA, I., AND MÜLLER, C. Location-scale depth. *Journal of the American Statistical Association* 99 (2004), 949–989.
12. R CORE TEAM. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2013.
13. ROMANO, C. Calibrating and simulating copula functions: An application to the Italian stock market. Tech. rep., Capitalia, November 2002.
14. SMITH, M. Modelling sample selection using Archimedean copulas. *Econometrics Journal* 6 (2003), 99–123.

Appendix 1: Calculating unbiased estimator of ρ from MLDE for t -copula

The true parameter value ρ is related to the MLDE $\hat{\rho}_d$ as follows:

$$\rho = \begin{cases} \text{sign}(\hat{\rho}_d) [c_3|\hat{\rho}_d|^3 + c_2|\hat{\rho}_d|^2 + c_1|\hat{\rho}_d| + c_0] & \text{if } |\hat{\rho}_d| > \rho_0 \\ 0 & \text{if } |\hat{\rho}_d| \leq \rho_0 \end{cases}$$

where the coefficients c_0, \dots, c_3 and the cutoff point ρ_0 are chosen from the row in Table 4 corresponding to the df of t -copula.

| df | c_0 | c_1 | c_2 | c_3 | ρ_0 |
|----|---------|--------|--------|--------|----------|
| 3 | -0.1272 | 0.5657 | 0.2685 | 0.3057 | 0.2272 |
| 4 | -0.1305 | 0.5061 | 0.3038 | 0.3343 | 0.2392 |
| 5 | -0.135 | 0.4826 | 0.2976 | 0.3686 | 0.2562 |
| 6 | -0.1419 | 0.5058 | 0.2062 | 0.4456 | 0.2693 |
| 7 | -0.1066 | 0.285 | 0.5389 | 0.2971 | 0.2853 |
| 8 | -0.1167 | 0.3246 | 0.4431 | 0.3649 | 0.2797 |
| 9 | -0.1075 | 0.2856 | 0.4717 | 0.3665 | 0.2881 |
| 10 | -0.1192 | 0.3207 | 0.4251 | 0.3889 | 0.2835 |
| 11 | -0.1575 | 0.5098 | 0.1136 | 0.5508 | 0.2956 |
| 12 | -0.146 | 0.4432 | 0.2057 | 0.5142 | 0.3079 |
| 13 | -0.1447 | 0.4199 | 0.2474 | 0.494 | 0.3038 |
| 14 | -0.1338 | 0.3706 | 0.3035 | 0.4758 | 0.3008 |
| 15 | -0.1334 | 0.3411 | 0.3674 | 0.4407 | 0.3186 |
| 16 | -0.1351 | 0.3739 | 0.2754 | 0.503 | 0.2964 |
| 17 | -0.1221 | 0.3006 | 0.3854 | 0.4526 | 0.2945 |
| 18 | -0.1106 | 0.2404 | 0.4772 | 0.4099 | 0.3101 |
| 19 | -0.1503 | 0.4428 | 0.1508 | 0.5744 | 0.2966 |
| 20 | -0.1636 | 0.5092 | 0.0411 | 0.6321 | 0.3135 |
| 21 | -0.1385 | 0.3805 | 0.2286 | 0.5483 | 0.3192 |
| 22 | -0.1429 | 0.3927 | 0.2227 | 0.5457 | 0.3042 |
| 23 | -0.1451 | 0.3991 | 0.2159 | 0.5477 | 0.3115 |
| 24 | -0.1265 | 0.3132 | 0.3345 | 0.4968 | 0.3179 |
| 25 | -0.136 | 0.3428 | 0.2983 | 0.5128 | 0.32 |
| 26 | -0.1141 | 0.2425 | 0.4325 | 0.4568 | 0.3115 |
| 27 | -0.1521 | 0.4334 | 0.1469 | 0.59 | 0.3127 |
| 28 | -0.1582 | 0.4675 | 0.0917 | 0.6171 | 0.3233 |
| 29 | -0.1521 | 0.4485 | 0.1003 | 0.6222 | 0.3176 |
| 30 | -0.1519 | 0.4504 | 0.0953 | 0.625 | 0.3119 |

Table 4. Table for t copula unbiased estimator calculation

Appendix 2: Description of variables in Cardiotocography data

| Variable name | Description (FHR = Fetal Heart Rate) |
|---------------|---|
| LB | FHR baseline (beats per minute) |
| AC | # of accelerations per second |
| FM | # of fetal movements per second |
| UC | # of uterine contractions per second |
| DL | # of light decelerations per second |
| DS | # of severe decelerations per second |
| DP | # of prolonged decelerations per second |
| ASTV | Percentage of time with abnormal short term variability |
| MSTV | Mean value of short term variability |
| ALTV | Percentage of time with abnormal long term variability |
| MLTV | Mean value of long term variability |
| Width | Width of FHR histogram |
| Min | Minimum of FHR histogram |
| Max | Maximum of FHR histogram |
| Nmax | # of histogram peaks |
| Nzeros | # of histogram zeros |
| Mode | Histogram mode |
| Mean | Histogram mean |
| Median | Histogram median |
| Variance | Histogram variance |
| Tendency | Histogram tendency |

Table 5. Description of variables in Cardiotocography data