

# Calibration of copulas

# Calibration of copulas

Goal: Determine a copula and the marginal distributions to model a given multi-dimensional data set.

## Calibration of copulas

Goal: Determine a copula and the marginal distributions to model a given multi-dimensional data set.

Input: A sample  $\{X_1, X_2, \dots, X_d\}$  of a c.d.f.  $F$  with continuous marginal distributions  $F_1, F_2, \dots, F_d$ .

## Calibration of copulas

Goal: Determine a copula and the marginal distributions to model a given multi-dimensional data set.

Input: A sample  $\{X_1, X_2, \dots, X_d\}$  of a c.d.f.  $F$  with continuous marginal distributions  $F_1, F_2, \dots, F_d$ .

Output: A copula  $C_\theta$  and an estimator  $\hat{\theta}$  for the parameter vector  $\theta$  of the copula  $C_\theta$  such which  $F(x) \approx C_{\hat{\theta}}(F_1(x_1), \dots, F_d(x_d))$  holds.

## Calibration of copulas

Goal: Determine a copula and the marginal distributions to model a given multi-dimensional data set.

Input: A sample  $\{X_1, X_2, \dots, X_d\}$  of a c.d.f.  $F$  with continuous marginal distributions  $F_1, F_2, \dots, F_d$ .

Output: A copula  $C_\theta$  and an estimator  $\hat{\theta}$  for the parameter vector  $\theta$  of the copula  $C_\theta$  such which  $F(x) \approx C_{\hat{\theta}}(F_1(x_1), \dots, F_d(x_d))$  holds.

Question 1: Which family of (known) copulas to use?

## Calibration of copulas

Goal: Determine a copula and the marginal distributions to model a given multi-dimensional data set.

Input: A sample  $\{X_1, X_2, \dots, X_d\}$  of a c.d.f.  $F$  with continuous marginal distributions  $F_1, F_2, \dots, F_d$ .

Output: A copula  $C_\theta$  and an estimator  $\hat{\theta}$  for the parameter vector  $\theta$  of the copula  $C_\theta$  such which  $F(x) \approx C_{\hat{\theta}}(F_1(x_1), \dots, F_d(x_d))$  holds.

Question 1: Which family of (known) copulas to use?

Answer: Selection of a suitable family of copulas based on (a) the visual comparison of the graphical representations of the data set on one side and of known copulas on the other, and (b) the empirical tail dependence coefficients.

## Calibration of copulas

Goal: Determine a copula and the marginal distributions to model a given multi-dimensional data set.

Input: A sample  $\{X_1, X_2, \dots, X_d\}$  of a c.d.f.  $F$  with continuous marginal distributions  $F_1, F_2, \dots, F_d$ .

Output: A copula  $C_\theta$  and an estimator  $\hat{\theta}$  for the parameter vector  $\theta$  of the copula  $C_\theta$  such which  $F(x) \approx C_{\hat{\theta}}(F_1(x_1), \dots, F_d(x_d))$  holds.

Question 1: Which family of (known) copulas to use?

Answer: Selection of a suitable family of copulas based on (a) the visual comparison of the graphical representations of the data set on one side and of known copulas on the other, and (b) the empirical tail dependence coefficients.

Question 2: What are the parameters of the prespecified family of copulas used for the modelling?

## Parameter estimation for $C_R^{Ga}$ , $C_{\nu,R}^t$ , $C_\theta^{Cl}$ and $C_\theta^{Gu}$

$$C_R^{Ga} = \phi_R^d(\phi^{-1}(u_1), \dots, \phi^{-1}(u_d))$$

$$R_{ij} = \sin(\pi(\rho_\tau)_{ij}/2)$$

## Parameter estimation for $C_R^{Ga}$ , $C_{\nu,R}^t$ , $C_\theta^{Cl}$ and $C_\theta^{Gu}$

$$C_R^{Ga} = \phi_R^d(\phi^{-1}(u_1), \dots, \phi^{-1}(u_d))$$

$$R_{ij} = \sin(\pi(\rho_\tau)_{ij}/2)$$

$$C_{\nu,R}^t = t_{\nu,R}^d(t_\nu^{-1}(u_1), \dots, t_\nu^{-1}(u_d))$$

$$R_{ij} = \sin(\pi(\rho_\tau)_{ij}/2)$$

## Parameter estimation for $C_R^{Ga}$ , $C_{\nu,R}^t$ , $C_\theta^{Cl}$ and $C_\theta^{Gu}$

$$C_R^{Ga} = \phi_R^d(\phi^{-1}(u_1), \dots, \phi^{-1}(u_d))$$

$$R_{ij} = \sin(\pi(\rho_\tau)_{ij}/2)$$

$$C_{\nu,R}^t = t_{\nu,R}^d(t_\nu^{-1}(u_1), \dots, t_\nu^{-1}(u_d))$$

$$R_{ij} = \sin(\pi(\rho_\tau)_{ij}/2)$$

$$C_\theta^{Gu}(u) = \exp\left(-[(-\ln u_1)^\theta + \dots + (-\ln u_d)^\theta]^{1/\theta}\right) \quad \theta = 1/(1 - (\rho_\tau)_{ij})$$

## Parameter estimation for $C_R^{Ga}$ , $C_{\nu,R}^t$ , $C_\theta^{Cl}$ and $C_\theta^{Gu}$

$$C_R^{Ga} = \phi_R^d(\phi^{-1}(u_1), \dots, \phi^{-1}(u_d))$$

$$R_{ij} = \sin(\pi(\rho_\tau)_{ij}/2)$$

$$C_{\nu,R}^t = t_{\nu,R}^d(t_\nu^{-1}(u_1), \dots, t_\nu^{-1}(u_d))$$

$$R_{ij} = \sin(\pi(\rho_\tau)_{ij}/2)$$

$$C_\theta^{Gu}(u) = \exp(-[(-\ln u_1)^\theta + \dots + (-\ln u_d)^\theta]^{1/\theta})$$

$$\theta = 1/(1 - (\rho_\tau)_{ij})$$

$$C_\theta^{Cl}(u) = (u_1^{-\theta} + \dots + u_d^{-\theta} - d + 1)^{-1/\theta}$$

$$\theta = 2(\rho_\tau)_{ij}/(1 - (\rho_\tau)_{ij})$$

## Parameter estimation for $C_R^{Ga}$ , $C_{\nu,R}^t$ , $C_\theta^{Cl}$ and $C_\theta^{Gu}$

$$C_R^{Ga} = \phi_R^d(\phi^{-1}(u_1), \dots, \phi^{-1}(u_d))$$

$$R_{ij} = \sin(\pi(\rho_\tau)_{ij}/2)$$

$$C_{\nu,R}^t = t_{\nu,R}^d(t_\nu^{-1}(u_1), \dots, t_\nu^{-1}(u_d))$$

$$R_{ij} = \sin(\pi(\rho_\tau)_{ij}/2)$$

$$C_\theta^{Gu}(u) = \exp(-[(-\ln u_1)^\theta + \dots + (-\ln u_d)^\theta]^{1/\theta})$$

$$\theta = 1/(1 - (\rho_\tau)_{ij})$$

$$C_\theta^{Cl}(u) = (u_1^{-\theta} + \dots + u_d^{-\theta} - d + 1)^{-1/\theta}$$

$$\theta = 2(\rho_\tau)_{ij}/(1 - (\rho_\tau)_{ij})$$

where

$$(\rho_\tau)_{ij} = \rho_\tau(X_{k,i}, X_{k,j})$$

$$= P((X_{k,i} - X_{l,i})(X_{k,j} - X_{l,j}) > 0) - P((X_{k,i} - X_{l,i})(X_{k,j} - X_{l,j}) < 0)$$

$$= E(\text{sign}((X_{k,i} - X_{l,i})(X_{k,j} - X_{l,j}))).$$

## Parameter estimation for $C_R^{Ga}$ , $C_{\nu,R}^t$ , $C_\theta^{Cl}$ and $C_\theta^{Gu}$

$$C_R^{Ga} = \phi_R^d(\phi^{-1}(u_1), \dots, \phi^{-1}(u_d))$$

$$R_{ij} = \sin(\pi(\rho_\tau)_{ij}/2)$$

$$C_{\nu,R}^t = t_{\nu,R}^d(t_\nu^{-1}(u_1), \dots, t_\nu^{-1}(u_d))$$

$$R_{ij} = \sin(\pi(\rho_\tau)_{ij}/2)$$

$$C_\theta^{Gu}(u) = \exp(-[(-\ln u_1)^\theta + \dots + (-\ln u_d)^\theta]^{1/\theta}) \quad \theta = 1/(1 - (\rho_\tau)_{ij})$$

$$C_\theta^{Cl}(u) = (u_1^{-\theta} + \dots + u_d^{-\theta} - d + 1)^{-1/\theta}$$

$$\theta = 2(\rho_\tau)_{ij}/(1 - (\rho_\tau)_{ij})$$

where

$$\begin{aligned} (\rho_\tau)_{ij} &= \rho_\tau(X_{k,i}, X_{k,j}) \\ &= P((X_{k,i} - X_{l,i})(X_{k,j} - X_{l,j}) > 0) - P((X_{k,i} - X_{l,i})(X_{k,j} - X_{l,j}) < 0) \\ &= E(\text{sign}((X_{k,i} - X_{l,i})(X_{k,j} - X_{l,j}))). \end{aligned}$$

Standard empirical estimator of Kendalls Tau:

$$\widehat{\rho_\tau}_{ij} = \binom{n}{2}^{-1} \sum_{1 \leq k < l \leq n} \text{sign}((X_{k,i} - X_{l,i})(X_{k,j} - X_{l,j})).$$

# Calibration of the correlation matrix for Gaussian and $t$ -copulas

## Calibration of the correlation matrix for Gaussian and $t$ -copulas

It may happen that  $\hat{R} = (\hat{R}_{ij})$ ,  $\hat{R}_{ij} = \sin(\pi \widehat{\rho_{\tau}}_{ij}/2)$ , is not positive definite.

## Calibration of the correlation matrix for Gaussian and $t$ -copulas

It may happen that  $\hat{R} = (\hat{R}_{ij})$ ,  $\hat{R}_{ij} = \sin(\pi \widehat{\rho_{\tau}}_{ij}/2)$ , is not positive definite. Replace  $\hat{R}$  by a correlation matrix  $R^*$ , selected such the “distance” between  $R^*$  and  $\hat{R}$  is “small”.

## Calibration of the correlation matrix for Gaussian and $t$ -copulas

It may happen that  $\hat{R} = (\hat{R}_{ij})$ ,  $\hat{R}_{ij} = \sin(\pi \widehat{\rho_{\tau ij}}/2)$ , is not positive definite. Replace  $\hat{R}$  by a correlation matrix  $R^*$ , selected such the “distance” between  $R^*$  and  $\hat{R}$  is “small”.

### Eigenvalue approach (Rousseeuw and Molenberghs 1993)

- Compute the spectral decomposition  $\hat{R} = \Gamma \Lambda \Gamma^T$  of  $\hat{R}$ , where  $\Lambda$  is a diagonal matrix, containing the eigenvalues of  $\hat{R}$  on the diagonal, and  $\Gamma$  is an orthogonal matrix with the eigenvectors of  $\hat{R}$  in its columns.

## Calibration of the correlation matrix for Gaussian and $t$ -copulas

It may happen that  $\hat{R} = (\hat{R}_{ij})$ ,  $\hat{R}_{ij} = \sin(\pi \widehat{\rho_{\tau}}_{ij}/2)$ , is not positive definite. Replace  $\hat{R}$  by a correlation matrix  $R^*$ , selected such the “distance” between  $R^*$  and  $\hat{R}$  is “small”.

### Eigenvalue approach (Rousseeuw and Molenberghs 1993)

- ▶ Compute the spectral decomposition  $\hat{R} = \Gamma \Lambda \Gamma^T$  of  $\hat{R}$ , where  $\Lambda$  is a diagonal matrix, containing the eigenvalues of  $\hat{R}$  on the diagonal, and  $\Gamma$  is an orthogonal matrix with the eigenvectors of  $\hat{R}$  in its columns.
- ▶ Replace the negative eigenvalues in  $\Lambda$  by some small number  $\delta > 0$  to obtain  $\tilde{\Lambda}$ .

## Calibration of the correlation matrix for Gaussian and $t$ -copulas

It may happen that  $\hat{R} = (\hat{R}_{ij})$ ,  $\hat{R}_{ij} = \sin(\pi \widehat{\rho_{\tau ij}}/2)$ , is not positive definite. Replace  $\hat{R}$  by a correlation matrix  $R^*$ , selected such the “distance” between  $R^*$  and  $\hat{R}$  is “small”.

### Eigenvalue approach (Rousseeuw and Molenberghs 1993)

- ▶ Compute the spectral decomposition  $\hat{R} = \Gamma \Lambda \Gamma^T$  of  $\hat{R}$ , where  $\Lambda$  is a diagonal matrix, containing the eigenvalues of  $\hat{R}$  on the diagonal, and  $\Gamma$  is an orthogonal matrix with the eigenvectors of  $\hat{R}$  in its columns.
- ▶ Replace the negative eigenvalues in  $\Lambda$  by some small number  $\delta > 0$  to obtain  $\tilde{\Lambda}$ .
- ▶ Compute  $\tilde{R} = \Gamma \tilde{\Lambda} \Gamma^T$ .  $\tilde{R}$  is symmetric and positive definite but not necessarily a correlation matrix; the diagonal elements  $\hat{R}_{ii}$  might be unequal 1.

## Calibration of the correlation matrix for Gaussian and $t$ -copulas

It may happen that  $\hat{R} = (\hat{R}_{ij})$ ,  $\hat{R}_{ij} = \sin(\pi \widehat{\rho_{\tau ij}}/2)$ , is not positive definite. Replace  $\hat{R}$  by a correlation matrix  $R^*$ , selected such the “distance” between  $R^*$  and  $\hat{R}$  is “small”.

### Eigenvalue approach (Rousseeuw and Molenberghs 1993)

- ▶ Compute the spectral decomposition  $\hat{R} = \Gamma \Lambda \Gamma^T$  of  $\hat{R}$ , where  $\Lambda$  is a diagonal matrix, containing the eigenvalues of  $\hat{R}$  on the diagonal, and  $\Gamma$  is an orthogonal matrix with the eigenvectors of  $\hat{R}$  in its columns.
- ▶ Replace the negative eigenvalues in  $\Lambda$  by some small number  $\delta > 0$  to obtain  $\tilde{\Lambda}$ .
- ▶ Compute  $\tilde{R} = \Gamma \tilde{\Lambda} \Gamma^T$ .  $\tilde{R}$  is symmetric and positive definite but not necessarily a correlation matrix; the diagonal elements  $\tilde{R}_{ii}$  might be unequal 1.
- ▶ Set  $R^* := D \tilde{R} D$  where  $D$  is a diagonal matrix with  $D_{k,k} = 1/\sqrt{\tilde{R}_{k,k}}$ .

# Estimation of the number of the degrees of freedom $\nu$ for $t$ -copulas

## Estimation of the number of the degrees of freedom $\nu$ for $t$ -copulas

1. Let  $\hat{F}_1, \dots, \hat{F}_d$  be the estimated marginal distributions.

## Estimation of the number of the degrees of freedom $\nu$ for $t$ -copulas

1. Let  $\hat{F}_1, \dots, \hat{F}_d$  be the estimated marginal distributions.
2. Generate a pseudo-sample of the copula

$$\hat{U}_k = (\hat{U}_{k,1}, \hat{U}_{k,2}, \dots, \hat{U}_{k,d}) := \left( \hat{F}_1(X_{k,1}), \dots, \hat{F}_d(X_{k,d}) \right),$$

for  $k = 1, 2, \dots, n$  (see Genest und Rivest 1993).

## Estimation of the number of the degrees of freedom $\nu$ for $t$ -copulas

1. Let  $\hat{F}_1, \dots, \hat{F}_d$  be the estimated marginal distributions.
2. Generate a pseudo-sample of the copula

$$\hat{U}_k = (\hat{U}_{k,1}, \hat{U}_{k,2}, \dots, \hat{U}_{k,d}) := \left( \hat{F}_1(X_{k,1}), \dots, \hat{F}_d(X_{k,d}) \right),$$

for  $k = 1, 2, \dots, n$  (see Genest und Rivest 1993).

$\hat{F}_k$  can be generated by :

- ▶ a parametric estimation method;  
 $\hat{F}_k$  is assumed to be a certain parametric distribution and the parameter is estimated by a maximum likelihood (ML) approach

## Estimation of the number of the degrees of freedom $\nu$ for $t$ -copulas

1. Let  $\hat{F}_1, \dots, \hat{F}_d$  be the estimated marginal distributions.
2. Generate a pseudo-sample of the copula

$$\hat{U}_k = (\hat{U}_{k,1}, \hat{U}_{k,2}, \dots, \hat{U}_{k,d}) := \left( \hat{F}_1(X_{k,1}), \dots, \hat{F}_d(X_{k,d}) \right),$$

for  $k = 1, 2, \dots, n$  (see Genest und Rivest 1993).

$\hat{F}_k$  can be generated by :

- ▶ a parametric estimation method;  
 $\hat{F}_k$  is assumed to be a certain parametric distribution and the parameter is estimated by a maximum likelihood (ML) approach
- ▶ a non-parametric estimation method;  
 $\hat{F}_i$  is the empirical distribution function  $\hat{F}_i(x) = \frac{1}{n+1} \sum_{t=1}^n I_{\{X_{t,i} \leq x\}}$ ,  
 $1 \leq i \leq d$ .

## Estimation of the number of the degrees of freedom $\nu$ for $t$ -copulas (contd.)

## Estimation of the number of the degrees of freedom $\nu$ for $t$ -copulas (contd.)

Maximum likelihood estimator of  $\nu$ :  $\nu = \arg \max_{\xi} \ln L(\xi; \hat{U}_1, \hat{U}_2, \dots, \hat{U}_n)$

## Estimation of the number of the degrees of freedom $\nu$ for $t$ -copulas (contd.)

Maximum likelihood estimator of  $\nu$ :  $\nu = \arg \max_{\xi} \ln L(\xi; \hat{U}_1, \hat{U}_2, \dots, \hat{U}_n)$   
where

$$L(\xi; \hat{U}_1, \hat{U}_2, \dots, \hat{U}_n) = \prod_{k=1}^n c_{\xi, R}^t(\hat{U}_k)$$

and  $c_{\xi, R}^t$  is the density of the  $t$ -copula  $C_{\xi, R}^t$ .

This implies

$$\ln L(\xi; \hat{U}_1, \hat{U}_2, \dots, \hat{U}_n) = \sum_{k=1}^n \ln g_{\xi, R}(t_{\xi}^{-1}(\hat{U}_{k,1}), \dots, t_{\xi}^{-1}(\hat{U}_{k,d})) - \sum_{k=1}^n \sum_{j=1}^d \ln g_{\xi}(t_{\xi}^{-1}(\hat{U}_{k,j})),$$

## Estimation of the number of the degrees of freedom $\nu$ for $t$ -copulas (contd.)

Maximum likelihood estimator of  $\nu$ :  $\nu = \arg \max_{\xi} \ln L(\xi; \hat{U}_1, \hat{U}_2, \dots, \hat{U}_n)$   
where

$$L(\xi; \hat{U}_1, \hat{U}_2, \dots, \hat{U}_n) = \prod_{k=1}^n c_{\xi,R}^t(\hat{U}_k)$$

and  $c_{\xi,R}^t$  is the density of the  $t$ -copula  $C_{\xi,R}^t$ .

This implies

$$\ln L(\xi; \hat{U}_1, \hat{U}_2, \dots, \hat{U}_n) = \sum_{k=1}^n \ln g_{\xi,R}(t_{\xi}^{-1}(\hat{U}_{k,1}), \dots, t_{\xi}^{-1}(\hat{U}_{k,d})) - \sum_{k=1}^n \sum_{j=1}^d \ln g_{\xi}(t_{\xi}^{-1}(\hat{U}_{k,j})),$$

where  $g_{\xi,R}$  is the density of a  $d$ -dimensional standard  $t$ -distribution with distribution function  $t_{\xi,R}^d$  and  $g_{\xi}$  is the density of a univariate standard  $t$ -distribution with  $\xi$  degrees of freedom.