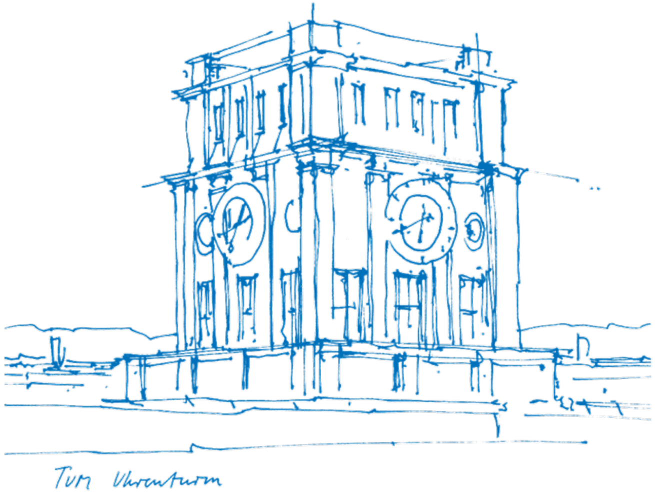
Statistical Modeling with Copulas

Aleksey Min

(based on the course materials of
Claudia Czado, Daniel Kraus and Matthias Scherer)

Technical University of Munich
Faculty of Mathematics

Winter term 2018/19



Statistical Modeling with Copulas

Acknowledgments

- Many thanks to:
  - Claudia Czado and Daniel Kraus for providing the course materials on Vine Copulas
  - Matthias Scherer for providing the course materials on Qunatitative Risk Management

Statistical Modeling with Copulas

Agenda

Lecture I: **Copulas, Sklar's theorem and ordinal measures of dependence**

Lecture II: **Archimedean and elliptical copulas**

Lecture III: **Estimation of copulas: parametric and semiparametric approaches**

Lecture IV: **Vine copulas**

Lecture V: **Estimation and model selection for vine copulas**

Lecture I

Copulas, Sklar's theorem and ordinal measures of dependence

Prominent example for copulas: The US mortgage crisis

Who uses portfolio credit-risk models?

- Investment banks
- Commercial banks
- Insurance companies
- Regulators
- Rating agencies

The US mortgage crisis

Dependence enters the game

The relevant task: understanding the **loss in the portfolio** up to $t > 0$

- This depends on:
  - The vector of **default times** $(\tau_1, \ldots, \tau_d) \in \mathbb{R}_+^d$
  - The **loss-given defaults** $1 - R_i$, where R_i is the recovery rate of firm i
  - The portfolio weights N_i
  - The **portfolio loss process** is then

$$\text{Loss}_t := \sum_{i=1}^{d} N_i \,(1 - R_i)\, \mathbf{1}_{\{\tau_i \leq t\}}, \quad t \geq 0$$

- The pricing of portfolio credit derivatives essentially requires

$$E[\, g(\text{Loss}_t)\,]$$

  for non-trivial functions g

The US mortgage crisis

[Li (2000)] “On Default Correlation: A copula function approach”

IIJ Institutional Investor Journals

DEMO JOURNALS PRACTICAL APPLICATIONS GUIDES E-COLLECTIONS AUTHOR INDEX REGI

The Journal of FIXED INCOME

Search:

CURRENT ISSUE | LIST OF ISSUES | JOURNAL INFORMATION | SUBMIT AN ARTICLE | EDIT

Article Abstract

Subscribe Now

On Default Correlation
A Copula Function Approach

The Journal of Fixed Income
March 2000, Vol. 9, No. 4: pp. 43-54
DOI: 10.3905/jfi.2000.319253

David X. Li

This article studies the problem of default correlation. It introduces a random variable called “time-until-default” to denote the survival time of each defaultable entity or financial instrument, and defines the default correlation between two credit risks as the correlation coefficient between their survival times. The author explains why a copula function approach should be used to specify the joint distribution of survival times after marginal distributions of survival times are derived from market information, such as risky bond prices or asset swap spreads. He shows that the current approach to default correlation through asset correlation is equivalent to using a normal copula function. Numerical examples illustrate the use of copula functions in the valuation of some credit derivatives, such as credit default swaps and first-to-default contracts.

http://dx.doi.org/10.3905/jfi.2000.319253

- Core content: Combine default times using a copula

Copulas and Sklar's Theorem

[Sklar (1959)]

- **Definition**

  $C : [0,1]^d \to [0,1]$ is called **copula**, if there is a random vector $(U_1, \dots, U_d)$ such that $U_k \sim \mathcal{U}[0,1]$ for each k and for $u_1, \dots, u_d \in [0,1]$:

$$C(u_1, \dots, u_d) = \mathbb{P}(U_1 \le u_1, \dots, U_d \le u_d)$$

- **Sklar's Theorem**

  $F : \mathbb{R}^d \to [0,1]$ is the distribution function of some random vector $(X_1, \dots, X_d)$ if and only if there exist a copula $C : [0,1]^d \to [0,1]$ and univariate distribution functions $F_1, \dots, F_d : \mathbb{R} \to [0,1]$ such that

$$C\big(F_1(x_1), \dots, F_d(x_d)\big) = F(x_1, \dots, x_d), \quad x_1, \dots, x_d \in \mathbb{R}$$

  The distribution of X_j equals F_j and the correspondence between F and C is one-to-one if all functions $F_1, \dots, F_d$ are continuous.

Copulas and Sklar's Theorem

The typical use in portfolio credit risk

Aim: Model a vector of default times

$$(\tau_1, \ldots, \tau_d)$$

(1) Fit marginal distribution functions

$$t \mapsto \mathbb{P}(\tau_k \leq t) =: F_k(t)$$

(2) Impose a (hopefully suitable) copula C on them to obtain the joint distribution F

$$F = C\big(F_1, \ldots, F_d\big)$$

Attention: This is mathematically valid, but is it reasonable?

Some statistical tools

Distribution functions and quantile functions

Definition 1 (Generalised inverse | Quantile function | α-Quantile)

(a) *Let h be an increasing (non-decreasing) function. With convention* $\inf \emptyset := \infty$, *one defines the* ***generalised inverse*** *of h as*

$$h^{\leftarrow} := \begin{cases} \mathbb{R} & \to \mathbb{R}, \\ y & \mapsto \inf\{x \in \mathbb{R} : h(x) \geq y\}. \end{cases}$$

(b) *Let* $F : \mathbb{R} \to [0, 1]$ *be the distribution function of a random variable* X, *i.e.* $F(x) := \mathbb{P}(X \leq x)$ *for all* $x \in \mathbb{R}$.

(i) *Then*

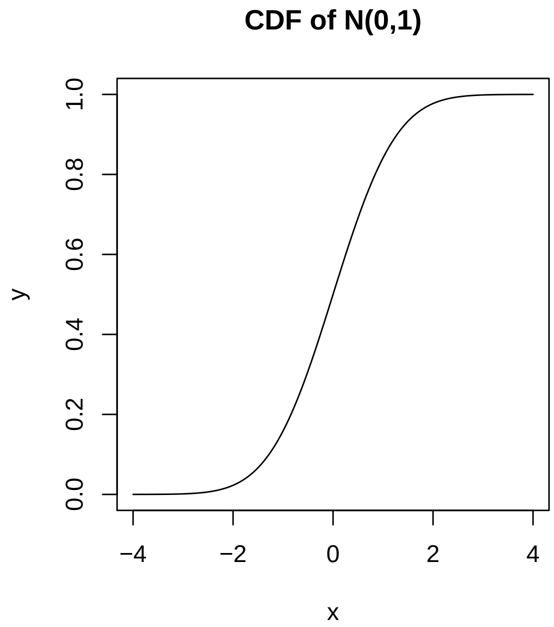
$$F^{\leftarrow} := \begin{cases} (0, 1) & \to \mathbb{R}, \\ y & \mapsto \inf\{x \in \mathbb{R} : F(x) \geq y\} \end{cases}$$

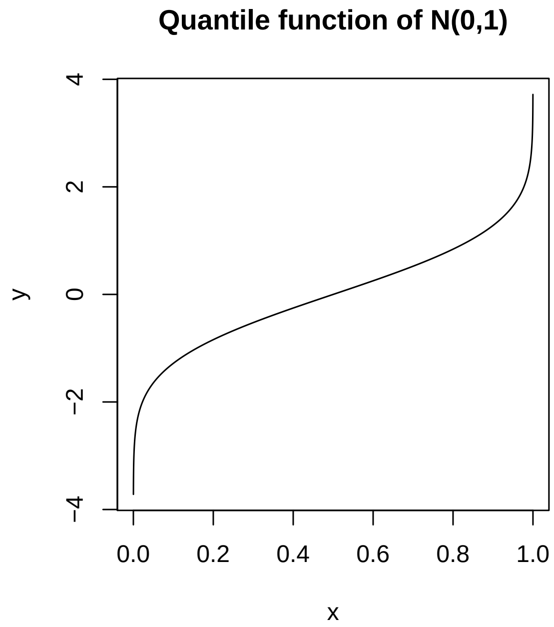
is the ***generalised inverse*** *or* ***quantile function****.*

(ii) *For* $\alpha \in (0, 1)$, *a number* $q_\alpha \in [F^{\leftarrow}(\alpha), F^{\leftarrow}(\alpha+)]$ *is called* α***-quantile*** *of* X.

Some statistical tools

Distribution functions and quantile functions

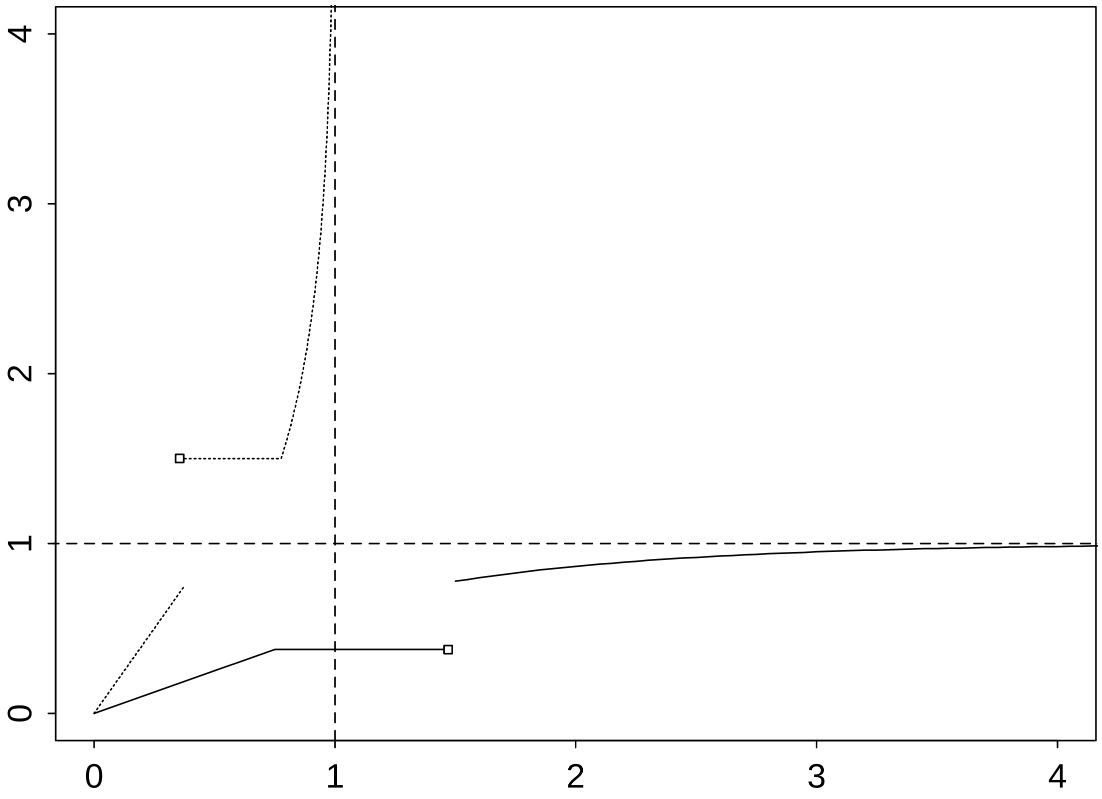




Distribution function of $X \sim \mathcal{N}(0, 1)$ (left) and its quantile function (right).

Some statistical tools

Distribution functions and quantile functions



A distribution function F (solid) and its quantile function $F^{\leftarrow}$ (dashed).

Some statistical tools

Distribution functions and quantile functions

Proposition 1

Let h be non-decreasing and right continuous. Denote by $h^{\leftarrow}$ its generalized inverse. The following properties hold:

(1) $h(x) \geq y \quad \Longleftrightarrow \quad x \geq h^{\leftarrow}(y)$.

(2) $h(x) < y \quad \Longleftrightarrow \quad x < h^{\leftarrow}(y)$.

(3) $h^{\leftarrow}$ *is non-decreasing and left-continuous.*

(4) $h \circ h^{\leftarrow}(y) \geq y$ *(with equality, if h is continuous).*

(5) $h^{\leftarrow} \circ h(x) \leq x$ *(with equality, if h is strictly increasing).*

(6) *h is strictly increasing* $\quad \Longleftrightarrow \quad$ $h^{\leftarrow}$ *is continuous.*

(7) *h is continuous* $\quad \Longleftrightarrow \quad$ $h^{\leftarrow}$ *is strictly increasing.*

(8) *h is strictly increasing and continuous on (a, b)* $\quad \Longrightarrow \quad h^{\leftarrow} = h^{-1}\big|_{(a,b)}$.

Some statistical tools

Distribution functions and quantile functions

Lemma 1

Let X be a random variable with distribution function F. Then

$$\mathbb{P}(F^{\leftarrow} \circ F(X) = X) = 1.$$

Proposition 2 (Probability integral transform)

Let X be a random variable with distribution function F and quantile function $F^{\leftarrow}$. Then:

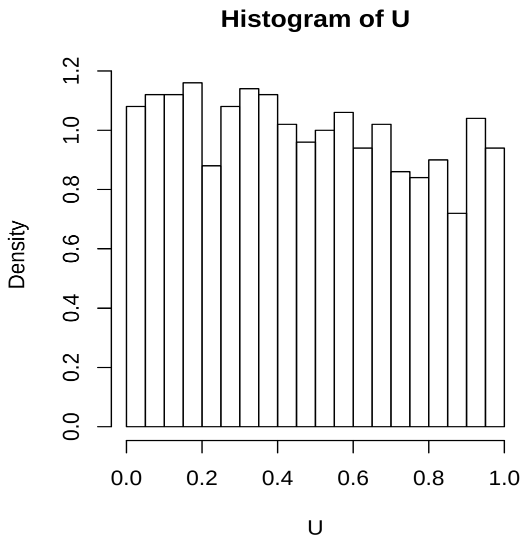
(1) *Let $U \sim \mathcal{U}(0,1)$, then $F^{\leftarrow}(U) \stackrel{d}{=} X$.*

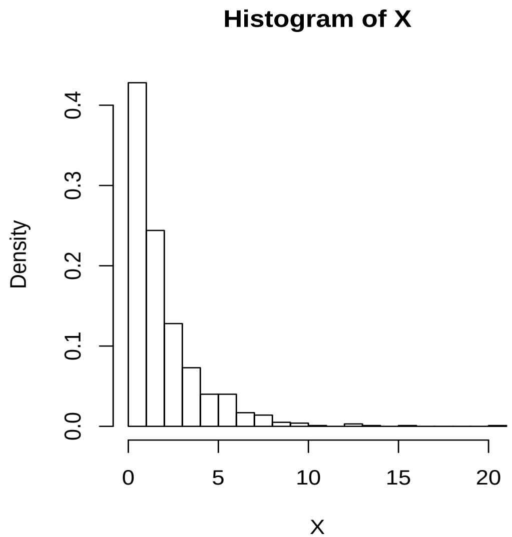
(2) $F(X) \stackrel{d}{=} U \quad \Longleftrightarrow \quad F$ *is continuous.*

Some statistical tools

Example in R: How Proposition 2 is used for sampling

```
n<-1000
U<-runif(n,0,1)
X<-(-1)*log(1-U)/0.5
hist(U,freq=FALSE,breaks=20)
hist(X,freq=FALSE,breaks=20)
```





Some statistical tools

Empirical distribution functions and quantile functions

Statistical methods are based on data

- In the simplest case, we have observations $x_1, \ldots, x_n$ that are real numbers.
- We consider each x_j as a realization of a random variable X (i.e. $x_j = X(\omega_j)$) for $j = 1, \ldots, n$, which are defined on the same probability space $(\Omega, \mathcal{F}, \mathbb{P})$.
- In the simplest case, these observations are independent and identically distributed (i.i.d.) and we want to estimate the distribution function F of X.

Some statistical tools

Empirical distribution functions and quantile functions

Definition 2 (Empirical distribution function | Empirical quantile function)

Let $X_1, \ldots, X_n$ be i.i.d. random variables with distribution function F and

$$X_{1,n} \leq X_{2,n} \leq \cdots \leq X_{n,n}$$

the corresponding order statistics. Then:

- *The **empirical distribution function** is given by*

$$F_n(x) = \frac{1}{n} \sum_{k=1}^{n} \mathbf{1}_{\{X_k \leq x\}}, \quad x \in \mathbb{R},$$

  i.e. $F_n(x) = k/n$ for $X_{k,n} \leq x < X_{k+1,n}$.

- *The **empirical quantile function** is given by*

$$F_n^{\leftarrow}(y) = \inf\{x \in \mathbb{R} : F_n(x) \geq y\} = X_{\lceil yn \rceil, n},$$

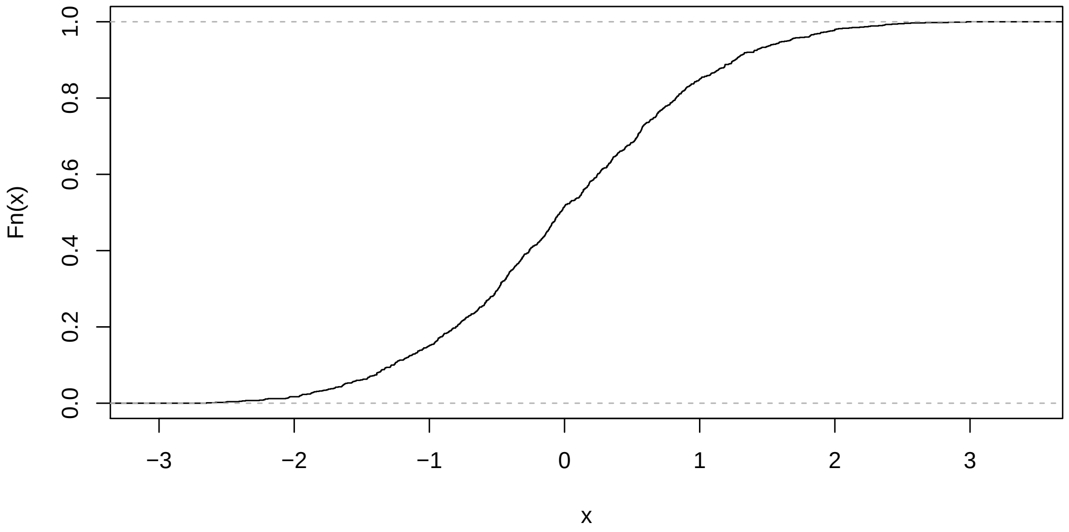
  where $\lceil z \rceil = \inf\{x \in \mathcal{Z} : z \leq x\}$.

For every $x \in \mathbb{R}$, $F_n(x)$ is a random variable and F_n is a random function.

Some statistical tools

Example in R: Empirical distribution function and Proposition 2

```
n<-1000
U<-runif(n,0,1)
X<-qnorm(U,0,1)
plot(ecdf(X), ylab="Fn(x)", verticals = FALSE, col.01line = "gray70", main="")
```



Some statistical tools

Empirical distribution functions and quantile functions

Theorem 1 (Glivenko–Cantelli)

Let $\{X_k\}_{k\in\mathbb{N}}$ be i.i.d. random variables on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ with distribution function F and empirical distribution function F_n. Then

$$\lim_{n\to\infty} \sup_{x\in\mathbb{R}} |F_n(x) - F(x)| = 0, \quad \mathbb{P}\text{-a.s.}$$

If F is strictly increasing, then for $y \in (0, 1)$

$$\lim_{n\to\infty} F_n^{\leftarrow}(y) = F^{\leftarrow}(y), \quad \mathbb{P}\text{-a.s.}$$

Some statistical tools

Empirical distribution functions and quantile functions

We now introduce two simple diagnostic tools to test, whether the observations $x_1, \ldots, x_n$ are realizations of i.i.d. random variables with distribution function $\widetilde{F}$.

Definition 3 (QQ-plot | PP-plot)

Let $x_1, \ldots, x_n$ be realizations of i.i.d. random variables with df F and

$$x_{1,n} \leq x_{2,n} \leq \cdots \leq x_{n,n}$$

the corresponding ordered values. Let $\widetilde{F}$ be some distribution function.

(1) *A **QQ-plot** consists of points $\{(\widetilde{F}^{\leftarrow}(\frac{k}{n+1}), x_{k,n})\}_{k=1,\ldots,n}$.*

(2) *A **PP-plot** consists of points $\{(\widetilde{F}(x_{k,n}), \frac{k}{n+1})\}_{k=1,\ldots,n}$.*

Some statistical tools

Empirical distribution functions and quantile functions

The interpretation of QQ-plots

- For risk management, concerning extreme events, a QQ-plot is more useful.
- Note that the first component of the QQ-plot is a theoretical quantile of $\widetilde{F}$ and the second the corresponding empirical quantile.
- More precisely, since

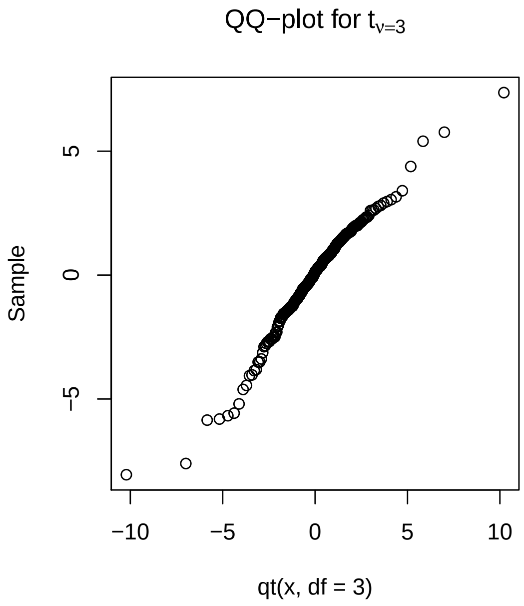
$$\frac{k-1}{n} < \frac{k}{n+1} < \frac{k}{n},$$

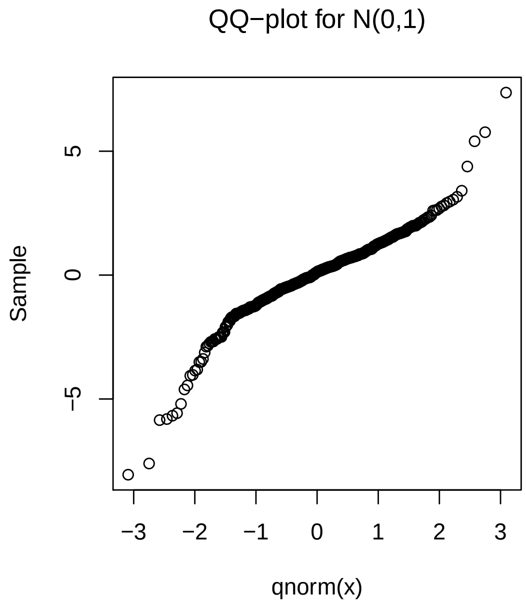
  and $F_n^{\leftarrow}(y) = x_{k,n}$ holds for $\frac{k-1}{n} < y < \frac{k}{n}$, we have $F_n^{\leftarrow}\left(\frac{k}{n+1}\right) = x_{k,n}$.
- Consequently, by Theorem 1 of Glivenko–Cantelli, if $\widetilde{F} \equiv F$, the points of the QQ-plot for large sample size n should lie approximately on the unit diagonal.

Some statistical tools

Example in R: QQ-plot

```
n<-500; Sample<-rt(n, df = 3); x<-ppoints(n);
qqplot(qt(x, df = 3), Sample, main = expression("QQ-plot for" ~ {t}[nu == 3]))
qqplot(qnorm(x), Sample, main = expression("QQ-plot for N(0,1)"))
```





Copulas and dependence structures

What can copulas do for you?

- They **describe and measure dependence** between random variables.
- They make it possible to **identify dependence**.
- They allow us to **construct new multivariate distributions**, with
  - arbitrary marginal laws,
  - all kinds of dependence structures.

$\Rightarrow$ Short: They separate marginal laws from dependence.

Copulas and dependence structures

What do you need to know for these tasks?

- A toolbox with **different copula families**.
- Understanding the analytical and **statistical properties** of different copulas.
- **Simulation** and **estimation** strategies.
- Understanding of **dependence measures**.

$\Rightarrow$ Short: This is only an introduction into a huge field.

Copulas and dependence structures

What are copulas?

Motivating example 1: Dependence between asset movements

- Consider three time series with daily observations (April 2008 to May 2013):
  - the stock price of BMW AG, $\{s_{t_i}^{(B)}\}_{i=0,1,2,\ldots,n}$,
  - the stock price of Daimler AG, $\{s_{t_i}^{(D)}\}_{i=0,1,2,\ldots,n}$,
  - a gold index, $\{s_{t_i}^{(G)}\}_{i=0,1,2,\ldots,n}$.
- Daily returns are defined as:

$$r_{t_{i+1}}^{(*)} := \frac{s_{t_{i+1}}^{(*)} - s_{t_i}^{(*)}}{s_{t_i}^{(*)}}, \quad i = 0, 1, 2, \ldots, n-1, \quad * \in \{B, D, G\}.$$

- Assume $r_{t_{i+1}}^{(*)}$, $i = 0, 1, 2, \ldots, n-1$, are i.i.d. samples from $R^{(*)}$, $* \in \{B, D, G\}$.
- We want to measure the dependence between $R^{(B)}$, $R^{(D)}$, and $R^{(G)}$.

Copulas and dependence structures

What are copulas?

Different methods for assessing dependence:

1. **Linear correlation:** The empirical (or historical) correlation coefficient of the BMW and gold index returns is given by

$$\hat{\rho}_n^{(B,G)} := \frac{\sum_{i=1}^{n} \left(r_{t_i}^{(B)} - \frac{1}{n}\sum_{j=1}^{n} r_{t_j}^{(B)} \right) \left(r_{t_i}^{(G)} - \frac{1}{n}\sum_{j=1}^{n} r_{t_j}^{(G)} \right)}{\sqrt{\sum_{i=1}^{n} \left(r_{t_i}^{(B)} - \frac{1}{n}\sum_{j=1}^{n} r_{t_j}^{(B)} \right)^2} \sqrt{\sum_{i=1}^{n} \left(r_{t_i}^{(G)} - \frac{1}{n}\sum_{j=1}^{n} r_{t_j}^{(G)} \right)^2}}.$$

   This is an estimator for Pearson's correlation coefficient ρ of $R^{(B)}$ and $R^{(G)}$

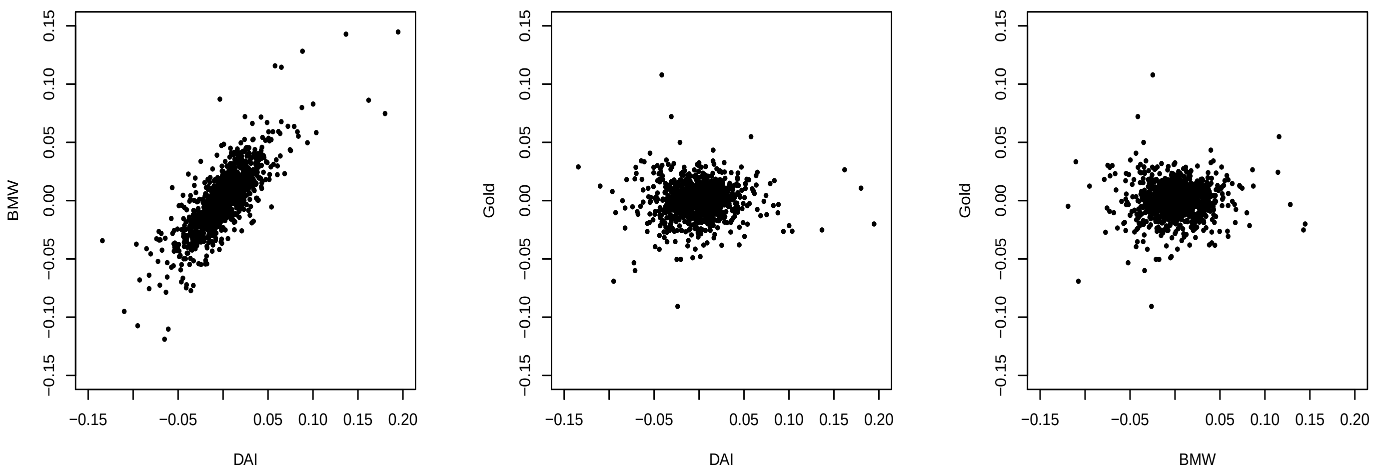
$$\rho^{(B,G)} := \mathrm{Cor}(R^{(B)}, R^{(G)}) := \frac{\mathbb{E}\left[(R^{(B)} - \mathbb{E}[R^{(B)}]) \, (R^{(G)} - \mathbb{E}[R^{(G)}]) \right]}{\sqrt{\mathbb{E}\left[(R^{(B)} - \mathbb{E}[R^{(B)}])^2 \right]} \sqrt{\mathbb{E}\left[(R^{(G)} - \mathbb{E}[R^{(G)}])^2 \right]}}.$$

   – It is the most popular dependence measure, although it measures only **linear** dependence.
   – In our example: $\hat{\rho}_n^{(B,D)} \gg \hat{\rho}_n^{(B,G)}$ and $\hat{\rho}_n^{(B,D)} \gg \hat{\rho}_n^{(D,G)}$ ($\approx 79.6\%$ vs. $\approx 4.4\%$).

Copulas and dependence structures

What are copulas?

2. **Scatter plot:** For each pair of $(R^{(B)}, R^{(D)}, R^{(G)})$, plot the observed historical data in a two-dimensional coordinate system:

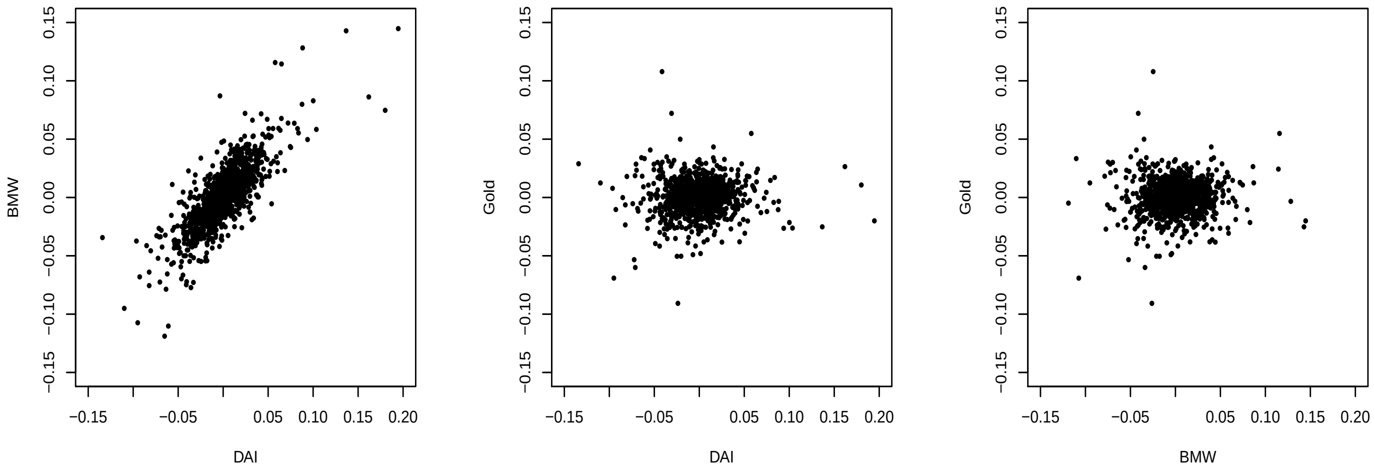


- All scatter plots are centered roughly around $(0, 0)$.
- The scatter plot of the two automobile firms is more elliptically shaped and more diagonal than the plot of gold vs. BMW or gold vs. Daimler.

Copulas and dependence structures

What are copulas?

3. **Concordance measurement:** A pair of points in the scatter plot is called **concordant**, if one point lies north east to the other one.

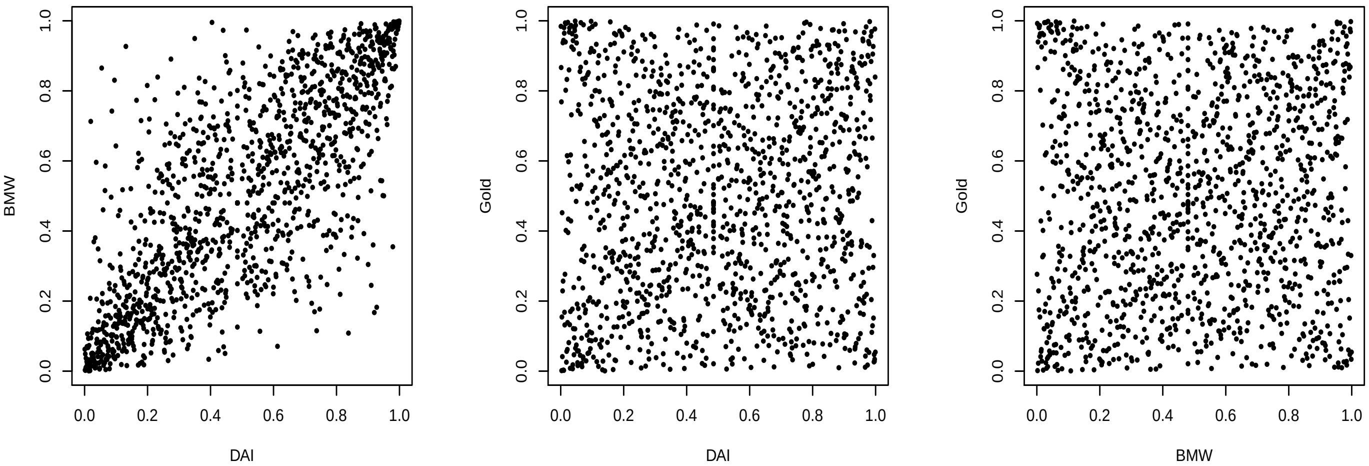


- Idea: Use only information about the *relative location* of the points.
- Concordance corresponds to positive dependence.
- Much more concordant pairs for BMW vs. Daimler than for gold vs. BMW or gold vs. Daimler.

Copulas and dependence structures

What are copulas?

3. **Rank transformation:** Replace each $r_{t_i}^{(*)}$ by its rank$/n$ within its time series $(r_{t_i}^{(*)})_{i=1,\ldots,n}$ and plot these new time series against each other.

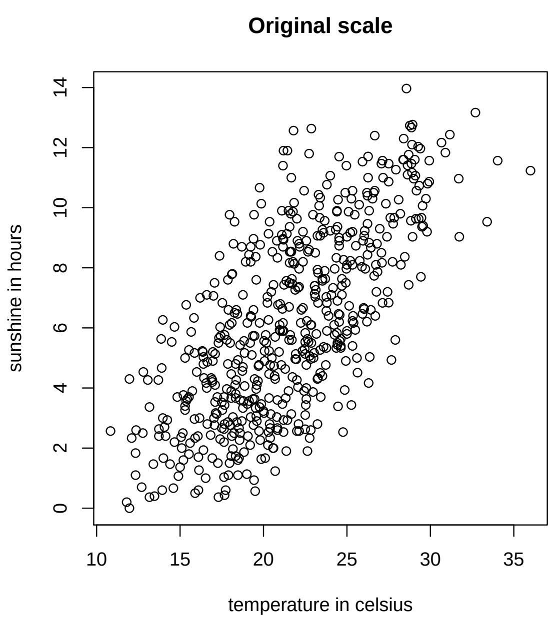


- Transformed time series live on $[0,1]$. Thus, the new scatter plot does not contain outliers like the original plot.
- The *dependence structure* remains unaltered: If two points in the original scatter plot were concordant, so are the newly assigned two points.

Copulas and dependence structures

What are copulas?

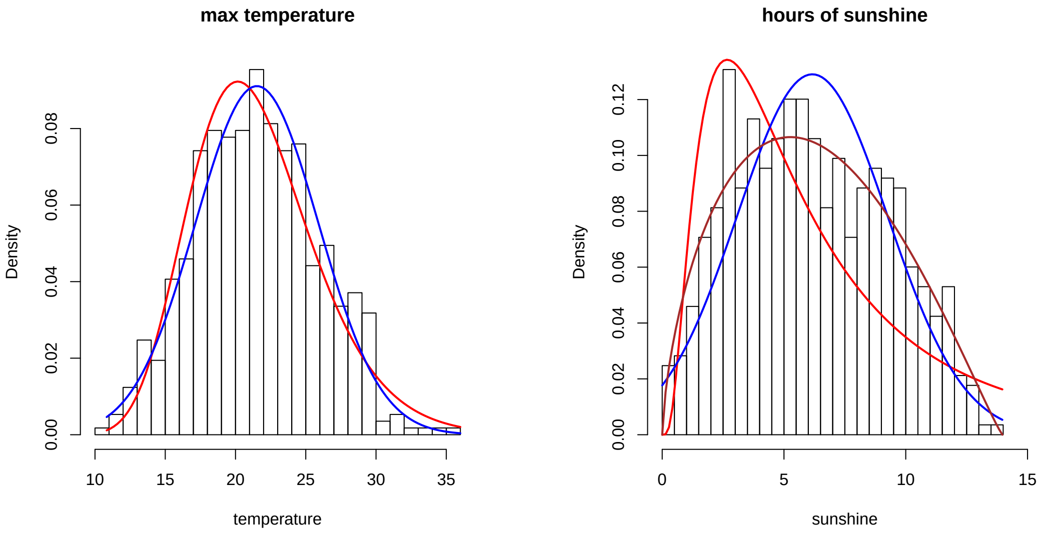
Motivating example 2: The “Biergarten” weather derivative



(Temperature, Sunshine Hours) of all weekends in August / September since 1948 in Regensburg.

Copulas and dependence structures

What are copulas?



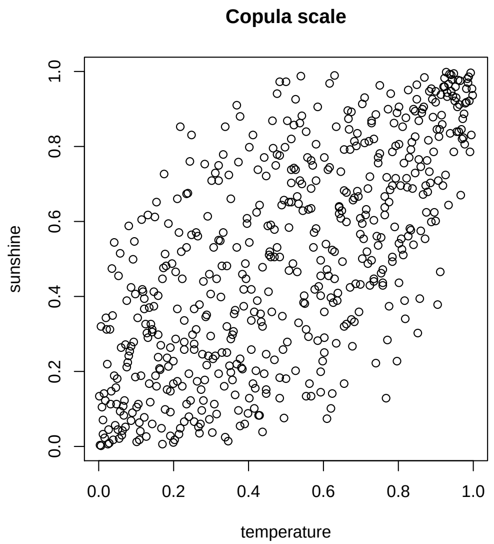
First consider the marginal laws:

Temperature F_1 is approximately normal.

Sunshine hours F_2 is approximately beta distributed with support $[0, 15]$.

Copulas and dependence structures

What are copulas?



- Rank transformation: Copula might be modeled with the Gaussian copula C.
- Hence, we can specify the joint distribution via $F(x_1, x_2) := C(F_1(x_1), F_2(x_2))$.

Copulas and dependence structures

Aim: Description of $(X_1, \ldots, X_d)$ on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$.

Definition 4 (Distribution function | margins)

- *The distribution function (d.f.) of $(X_1, \ldots, X_d)$ is defined as*

$$F(x_1, \ldots, x_d) := \mathbb{P}(X_1 \le x_1, \ldots, X_d \le x_d), \quad x_1, \ldots, x_d \in \mathbb{R}.$$

- *The one-dimensional distribution functions*

$$F_j(x) := \mathbb{P}(X_j \le x), \quad x \in \mathbb{R},$$

of the components X_j, $j = 1, \ldots, d$, are called "(one-dimensional) marginals" or "(one-dimensional) margins" of the d.f. of the random vector $(X_1, \ldots, X_d)$.

Remark 1

The distribution function F characterizes the probability law of $(X_1, \ldots, X_d)$.

Copulas and dependence structures

Definition 5 (Copula)

A function

$$C : [0, 1]^d \to [0, 1]$$

is called copula, if there is a random vector $(U_1, \ldots, U_d)$ *such that:*

a) *Each margin* U_j, $j = 1, \ldots, d$, *is uniform on* $[0, 1]$, *i.e.*

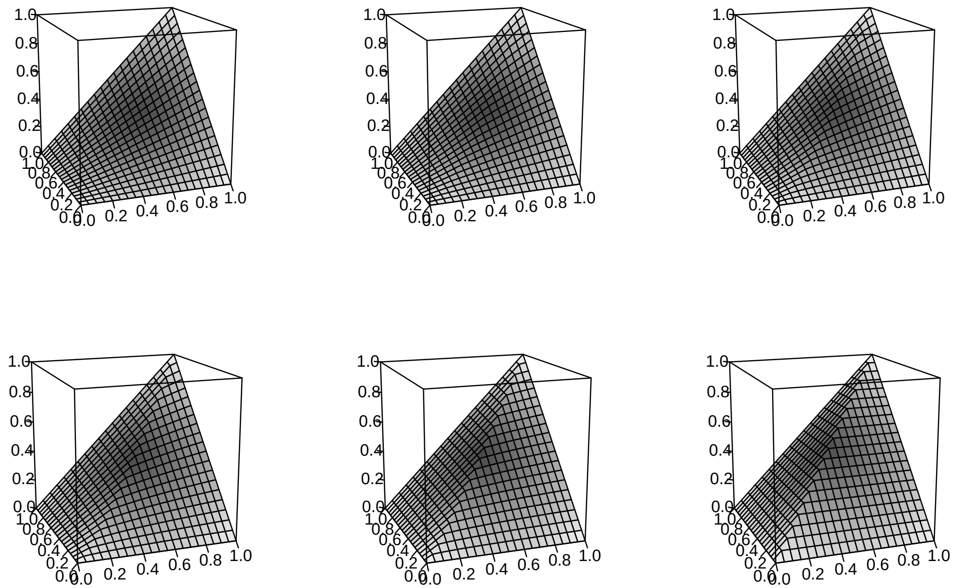
$$U_j \sim \mathcal{U}[0, 1],$$

b) C *is the joint distribution of* $(U_1, \ldots, U_d)$, *i.e.*

$$C(u_1, \ldots, u_d) = \mathbb{P}(U_1 \le u_1, \ldots, U_d \le u_d), \quad u_1, \ldots, u_d \in [0, 1].$$

Copulas and dependence structures

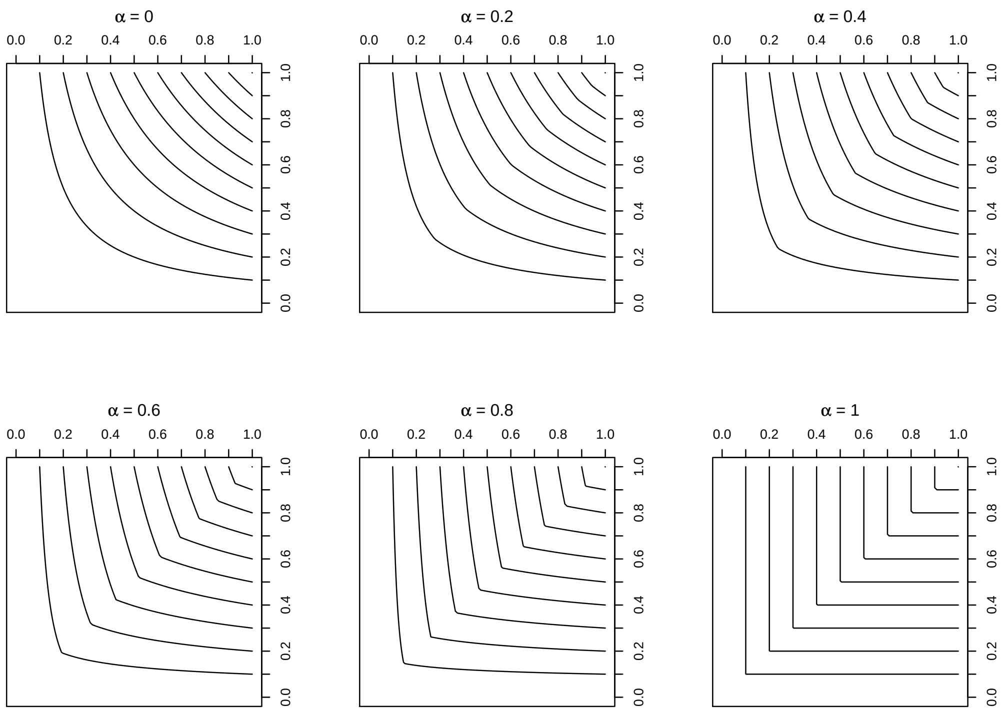
Graphical visualization as functions



Cuadras–Augé copula $C_\alpha(u_1, u_2) = \min\{u_1, u_2\} \max\{u_1, u_2\}^{1-\alpha}$,
$\alpha = 0$ (upper left), $\alpha = 0.2$, $\alpha = 0.4$, $\alpha = 0.6$, $\alpha = 0.8$, and $\alpha = 1$ (lower right).

Copulas and dependence structures

Graphical visualization as level plots

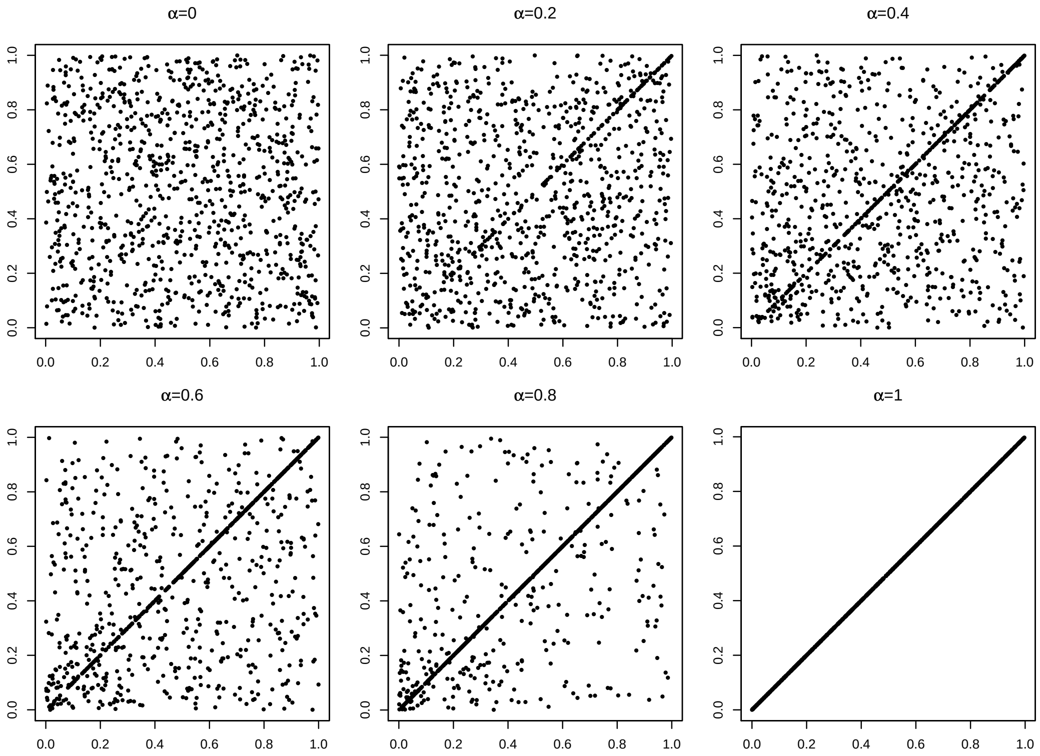


It is often more illustrative to plot a discrete grid of level sets

$$L_{k,n} := \left\{(u_1, u_2) \in [0,1]^2 \;:\; C(u_1, u_2) = k/n\right\}, \quad k = 0, 1, \ldots, n.$$

Copulas and dependence structures

Graphical visualization as scatter plots



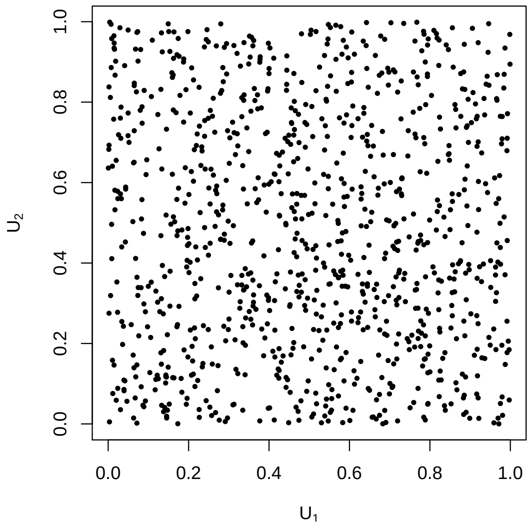
Interpretation: $\frac{\text{number of points in } B}{\text{number of all points}} \approx \mathbb{P}\big((U_1, U_2) \in B\big) = dC(B), \quad B \in \mathcal{B}(\mathbb{R}^2).$

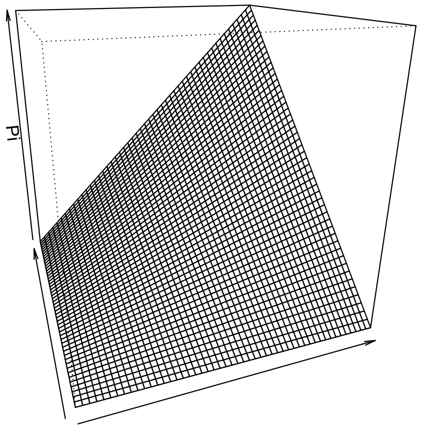
First examples

Independence copula:

- Consider d independent random variables $U_j \sim \mathcal{U}[0,1]$, $j = 1, 2, \ldots, d$.
- The joint distribution of $(U_1, U_2, \ldots, U_d)$ is the **independence copula**

$$\Pi_d(u_1, u_2, \ldots, u_d) := \prod_{j=1}^{d} u_j, \quad u_1, \ldots, u_d \in [0, 1].$$





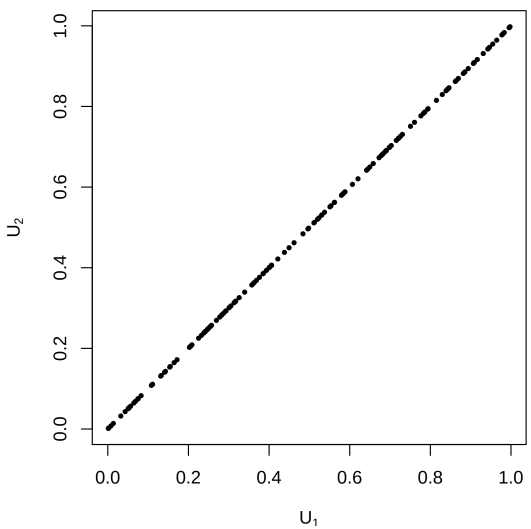
Copulas and dependence structures

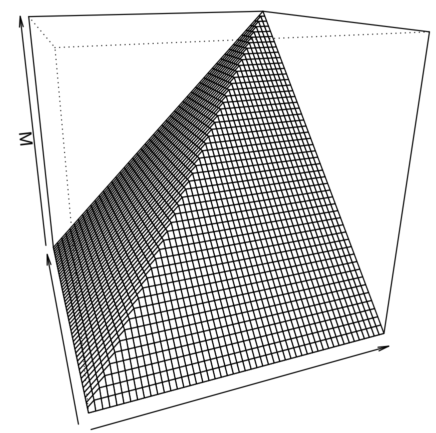
First examples

Comonotonicity copula:

- Consider for $U \sim \mathcal{U}[0,1]$ the vector $(U,\ldots,U)$.
- The joint distribution of $(U,\ldots,U)$ is the **comonotonicity copula**

$$M_d(u_1,\ldots,u_d) := \min\{u_1,\ldots,u_d\}, \quad u_1,\ldots,u_d \in [0,1].$$





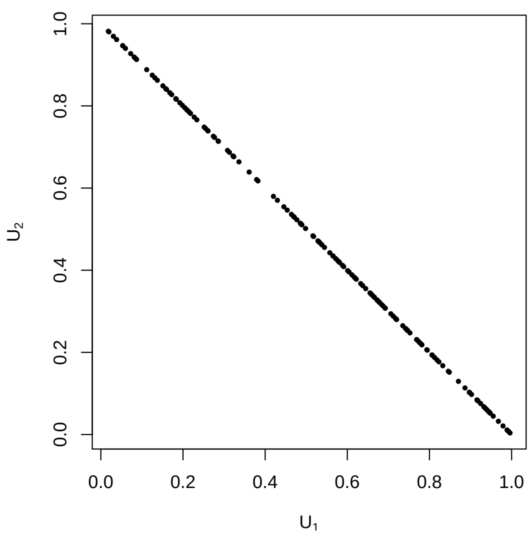
Copulas and dependence structures

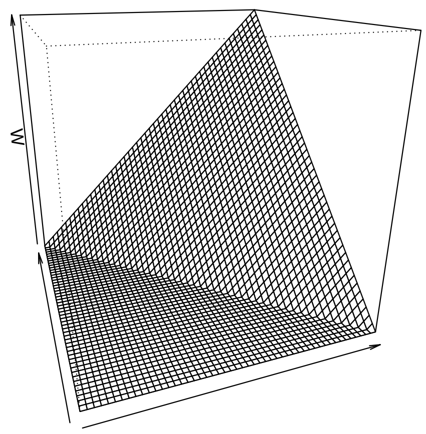
First examples

Countermonotonicity copula:

- Consider for $U \sim \mathcal{U}[0,1]$ the vector $(U, 1-U)$.
- The joint distribution of $(U, 1-U)$ is the **countermonotonicity copula**

$$W_2(u_1, u_2) := (u_1 + u_2 - 1)\,\mathbf{1}_{\{u_1+u_2 \geq 1\}}, \quad u_1, u_2 \in [0,1].$$





Copulas and dependence structures

Remark 2 (Alternative definition of a copula)

A copula is a function $C : [0,1]^d \to [0,1]$ *that satisfies the properties:*

(i) ***Groundedness****: Whenever at least one argument* $u_j = 0$*, then* $C(u_1, \ldots, u_d) = 0$*. This reflects*

$$0 \leq \mathbb{P}(U_1 \leq u_1, \ldots, U_j \leq 0, \ldots, U_d \leq u_d) \leq \mathbb{P}(U_j \leq 0) = 0.$$

(ii) ***Normalized marginals****:* $C(1, \ldots, 1, u_j, 1 \ldots, 1) = u_j$*, for* $u_j \in [0,1]$*. This reflects the uniform marginals property, since*

$$\mathbb{P}(U_1 \leq 1, \ldots, U_j \leq u_j, \ldots, U_d \leq 1) = \mathbb{P}(U_j \leq u_j) = u_j.$$

(iii) d***-increasingness****: For each* d*-dimensional rectangle* $\times_{j=1}^{d}[a_j, b_j]$*, being a subset of* $[0,1]^d$*, one has:*

$$0 \leq \sum_{(c_1,\ldots,c_d) \in \times_{j=1}^{d}\{a_j, b_j\}} (-1)^{|\{j : c_j = a_j\}|} C(c_1, \ldots, c_d) \leq 1.$$

Copulas and dependence structures

Sklar's Theorem

Theorem 2 (Sklar's Theorem)

A function $F : \mathbb{R}^d \to [0,1]$ is the distribution function of some random vector $(X_1, \dots, X_d)$ if and only if there exist a copula $C : [0,1]^d \to [0,1]$ and univariate distribution functions $F_1, \dots, F_d : \mathbb{R} \to [0,1]$, such that

$$C\big(F_1(x_1), \dots, F_d(x_d)\big) = F(x_1, \dots, x_d), \quad x_1, \dots, x_d \in \mathbb{R}. \tag{1}$$

The distribution function of component X_j equals F_j, $j = 1, \dots, d$, and the link between F and C is one-to-one if all functions $F_1, \dots, F_d$ are continuous.

Remark 3

Sklar's Theorem allows to subdivide the handling of the probability law of a random vector $(X_1, \dots, X_d)$ into two subsequent tasks:

1. *Handling of the one-dimensional marginal distribution functions.*
2. *Handling of the isolated dependence structure in the form of a copula.*

Copulas and dependence structures

Sklar's Theorem

Remark 4 (Uniqueness of the copula fails for non-continuous margins)

If the marginals $F_1, \ldots, F_d$ are not continuous, then there exist at least two copulas $C_1 \neq C_2$ both satisfying

$$C_1\big(F_1(x_1), \ldots, F_d(x_d)\big) = F(x_1, \ldots, x_d) = C_2\big(F_1(x_1), \ldots, F_d(x_d)\big)$$

for all $x_1, \ldots, x_d \in \mathbb{R}$. In most financial applications of copulas the margins are continuous, so this ambiguity is not an issue.

Copulas and dependence structures

Sklar's Theorem

Corollary 1

Take the notations from Theorem 2 and assume the marginals are continuous. Then, for any random vector $(X_1, \ldots, X_d) \sim F$*, we have*

$$(U_1, \ldots, U_d) := \big(F_1(X_1), \ldots, F_d(X_d)\big) \sim C. \tag{2}$$

On the other hand, for any random vector $(U_1, \ldots, U_d) \sim C$*, it holds*

$$(X_1, \ldots, X_d) := \big(F_1^{-1}(U_1), \ldots, F_d^{-1}(U_d)\big) \sim F, \tag{3}$$

and

$$C(u_1, \ldots, u_d) = F\big(F_1^{-1}(u_1), \ldots, F_d^{-1}(u_d)\big), \quad u_1, \ldots, u_d \in (0, 1).$$

Remark: In this chapter, generalized inverses are denote by F_i^{-1}.

Copulas and dependence structures

Sklar's Theorem

Sklar's Theorem can be applied in two directions:

(a) Analyzing distribution functions:

$$F \rightsquigarrow C \oplus (F_1, \ldots, F_d).$$

 (i) Analyze the univariate marginals (i.e. $F_1, \ldots, F_d$), using either a parametric or a nonparametric approach.
 (ii) Analyze the remaining dependence (i.e. C).

(b) Constructing distribution functions:

$$C \oplus (F_1, \ldots, F_d) \rightsquigarrow F.$$

Copulas and dependence structures

Sklar's Theorem

(a) Analyzing distribution functions – the marginals

(i) **Parametric approach**:

- **Assumption**: Marginal F_j stems from a certain parametric family of distribution functions, e.g. $F_j(x) = 1 - \exp(-\lambda_j x)$, $x \geq 0$.
- **Aim**: Estimate the unknown parameter(s), e.g. the parameter $\lambda_j > 0$.
- **Advantage**: Estimation routines for the parameters are known for many popular parametric families, e.g. in the exponential case $\hat{\lambda}_{j,n} = n / \sum_{i=1}^{n} X_j^{(i)}$. The fitted model can be used in all further investigations, e.g. the estimation of the dependence structure.
- **Disadvantage**: The observed data might not be explained very good by any member of the assumed parametric family (i.e. model risk).

Copulas and dependence structures

Sklar's Theorem

(a) Analyzing distribution functions – the marginals

(ii) **Non-parametric approach**:

- **Advantage**: The whole function $x \mapsto F_j(x)$ is estimated from the data, no (or only a little) pre-knowledge is needed.
- **Example**: "*Empirical distribution function*", which is defined by

$$\hat{F}_{j,n}(x) := \frac{1}{n} \sum_{i=1}^{n} \mathbf{1}_{\{X_j^{(i)} \leq x\}}, \quad x \in \mathbb{R},$$

which is well-known to converge almost surely and uniformly in x to the true distribution function F_j of X_j, as $n \to \infty$.

Copulas and dependence structures

Sklar's Theorem

(a) Analyzing distribution functions – the copula

- Given: Estimated marginals $\hat{F}_{1,n}, \ldots, \hat{F}_{d,n}$.
- Recall $(U_1, \ldots, U_d) := \big(F_1(X_1), \ldots, F_d(X_d)\big) \sim C$ if the margins are continuous.
- The random vectors

$$\big(\widehat{U}_1^{(i)}, \ldots, \widehat{U}_d^{(i)}\big) := \big(\hat{F}_{1,n}(X_1^{(i)}), \ldots, \hat{F}_{d,n}(X_d^{(i)})\big), \quad i = 1, \ldots, n,$$

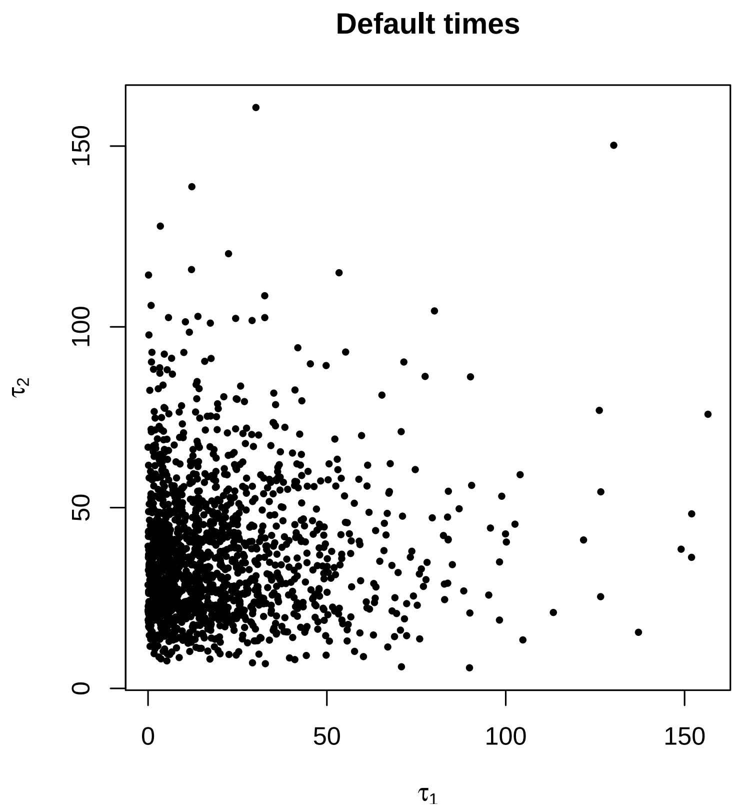
  are called "pseudo-observations".
- Estimate the copula C based on these samples:

(i) Parametric approach, e.g. compute empirical counterparts to copula-based dependence measures or use maximum-likelihood.

(ii) Non-parametric approach, e.g. multivariate empirical distribution function.

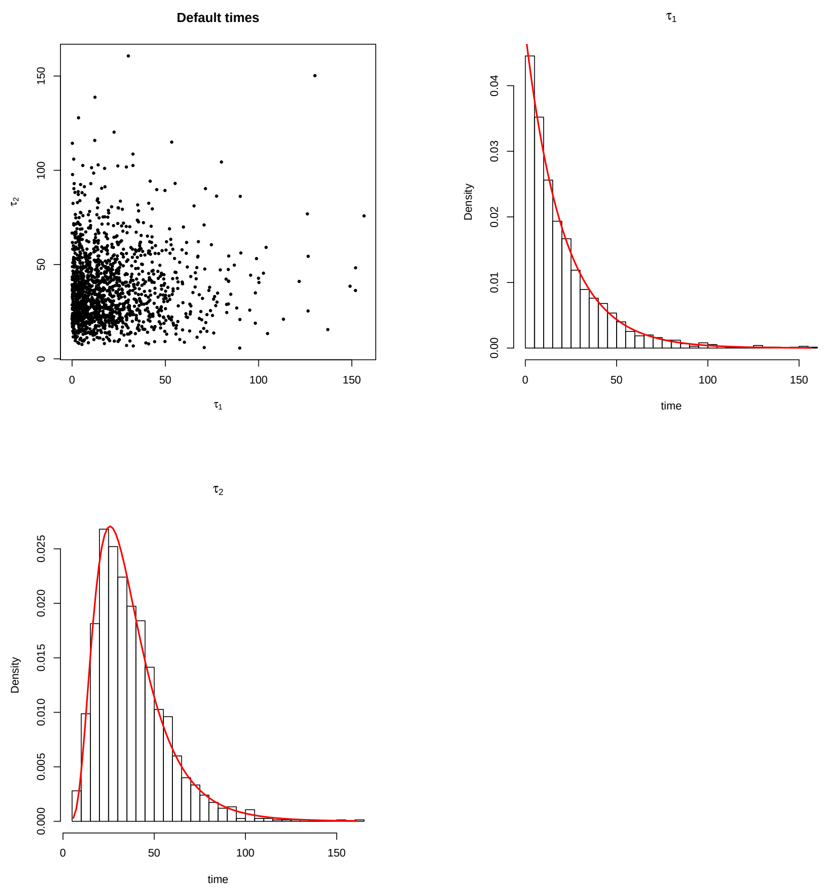
Copulas and dependence structures

Example: Realizations of (τ_1, τ_2). Can you guess the dependence?



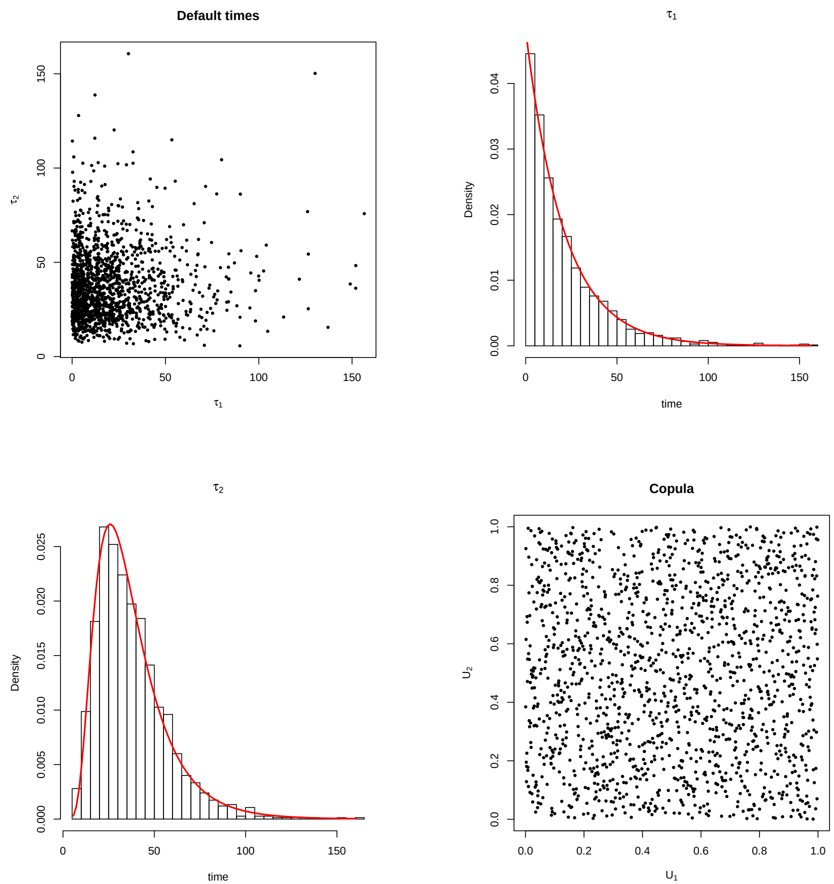
Copulas and dependence structures

Step 1: Identify the marginals



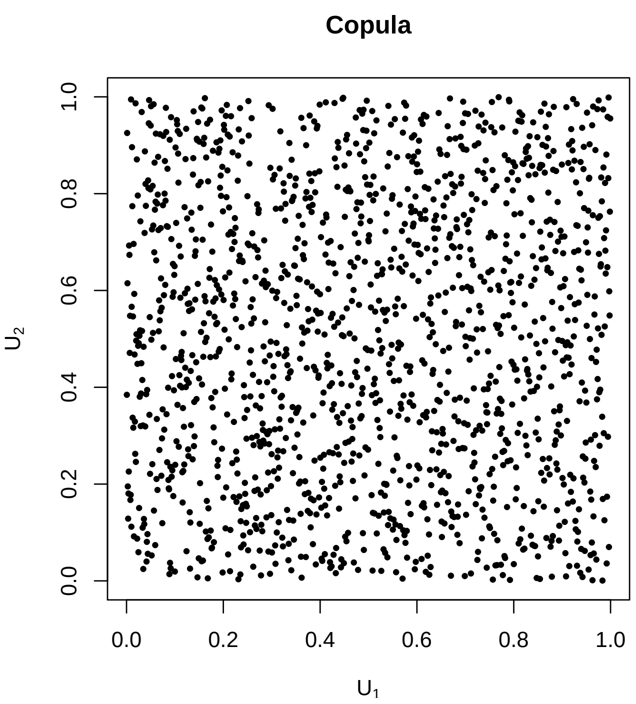
Copulas and dependence structures

Step 2: Transform marginals to pseudo-observations on $[0, 1]^2$



Copulas and dependence structures

Surprise, it actually was independence!



Influence of margins makes it difficult to identify the dependence!

Copulas and dependence structures

Sklar's Theorem

(b) Constructing distribution functions

- **Given**: Copula C and univariate distribution functions $F_1, \ldots, F_d$.
- **Aim**: Multivariate distribution function F.
- Think of situations when:
  - There is good knowledge about the single components (i.e. $F_1, \ldots, F_d$) but
  - little knowledge about the dependence structure (i.e. C).
  - Provided only a few observations, a high-dimensional model must be inferred (e.g. portfolio credit-risk modeling).
- **Approach**: Choose C from some flexible, parametric family of copulas, and fit the parameter(s) to the limited observable data.
- **Problem**: Lots of assumptions on the underlying copula are necessary.

Basic rules for working with copulas

Fréchet–Hoeffding bounds

Definition 6 (Fréchet–Hoeffding bounds)

The Fréchet–Hoeffding bounds for a $d-$dimensional copula are defined as

$$W_d(u_1,\ldots,u_d) := \max\{u_1+\ldots+u_d-(d-1),0\} \qquad \text{(“lower Fréchet–Hoeffding bound”)},$$
$$M_d(u_1,\ldots,u_d) := \min\{u_1,\ldots,u_d\} \qquad \text{(“upper Fréchet–Hoeffding bound”)}.$$

Theorem 3 (Fréchet–Hoeffding bounds)

Let $C:[0,1]^d \to [0,1]$ be an arbitrary copula. Then C is bounded by

$$W_d(u_1,\ldots,u_d) \le C(u_1,\ldots,u_d) \le M_d(u_1,\ldots,u_d), \quad u_1,\ldots,u_d \in [0,1].$$

These bounds are sharp in the sense that M_d is itself a copula, and for each point $\mathbf{u} := (u_1,\ldots,u_d) \in [0,1]^d$ one can find a copula $C_\mathbf{u}$ satisfying the equality

$$C_\mathbf{u}(u_1,\ldots,u_d) = W_d(u_1,\ldots,u_d).$$

Basic rules for working with copulas

Fréchet–Hoeffding bounds

Remark 5

1. *The Fréchet–Hoeffding bounds might be viewed as the extreme cases of most negative and most positive dependence.*
2. *A random vector* $(U_1, \ldots, U_d)$ *has* M_d *as joint distribution function if and only if* $U_1 = \ldots = U_d$ *holds with probability one,* M_d *is the "comonotonicity copula".*
3. W_d *is a copula only for* $d = 2$ *("countermonotonicity copula").* (U_1, U_2) *has* W_2 *as joint distribution function if and only if* $U_1 = 1 - U_2$ *holds with probability one.*

Basic rules for working with copulas

Fréchet–Hoeffding bounds

“Middle case” of stochastic independence: Unlike a linear correlation coefficient of 0, the “*independence copula*” or “*product copula*”

$$\Pi_d(u_1, \ldots, u_d) = u_1 \cdot u_2 \cdots u_d$$

really means stochastic independence.

Lemma 2 (Independence $\Leftrightarrow C = \Pi_d$)

A random vector $(X_1, \ldots, X_d)$ has stochastically independent components if and only if its distribution function can be split into its marginals and the copula Π_d, i.e.

$$\begin{aligned} F(x_1, \ldots, x_d) &= \mathbb{P}(X_1 \leq x_1, \ldots, X_d \leq x_d) \\ &= \mathbb{P}(X_1 \leq x_1) \cdot \ldots \cdot \mathbb{P}(X_d \leq x_d) \\ &= \Pi_d\big(F_1(x_1), \ldots, F_d(x_d)\big). \end{aligned}$$

Basic rules for working with copulas

Invariance under strictly monotone transformations

Recall: Transforming the components of a random vector $(X_1, \ldots, X_d)$ changes its distribution function. However, the dependence structure is not affected by strictly monotone transformations.

Lemma 3 (Strictly monotone transformations)

Let $(X_1, \ldots, X_d)$ *be a random vector with continuous marginals and copula* C*. For functions* $g_1, \ldots, g_d : \mathbb{R} \to \mathbb{R}$*, which are strictly increasing on the range of the respective components, the copula of* $\big(g_1(X_1), \ldots, g_d(X_d)\big)$ *is again* C*.*

Basic rules for working with copulas

Invariance under strictly monotone transformations

Remark 6 (Where is this useful?)

This allows to change marginals of a random vector at one's personal taste:

- *Let* $(X_1, \ldots, X_d)$ *have strictly increasing and continuous marginals* $F_1, \ldots, F_d$.
- *Let* $\tilde{F}_1, \ldots, \tilde{F}_d$ *be strictly increasing and continuous distribution functions.*
- *Define* $(\tilde{X}_1, \ldots, \tilde{X}_d)$ *by* $\tilde{X}_i = \tilde{F}_i^{-1} \circ F_i(X_i)$, *such that* $\tilde{X}_i \sim \tilde{F}_i$.
- *Lemma 3 shows that the copula is not affected by such a transformation.*
- *An example is the "probability integral transformation"*

$$(U_1, \ldots, U_d) := \big(F_1(X_1), \ldots, F_d(X_d)\big)$$

that standardizes the margins to uniform distributions on $[0, 1]$.

Basic rules for working with copulas

Invariance under strictly monotone transformations

Example 1 (Where is this useful in practice?)

For the dependence structure (i.e. copula) it does not matter whether one . . .

- *looks at values of stock prices or at their logarithmic values,*
- *converts prices in other currencies by multiplication with FX rates,*
- *changes the scale of credit spreads from percent into basis points.*

This invariance of the copula under strictly increasing margin transformations is not shared by the popular concept of correlation coefficients!

Basic rules for working with copulas

Invariance under strictly monotone transformations

Corollary 2 ($C = M_d \Leftrightarrow$ comonotonicity)

A random vector $(X_1, \ldots, X_d)$ *with marginals* $F_1, \ldots, F_d$ *has copula* M_d *if and only if*

$$(X_1, \ldots, X_d) \stackrel{d}{=} \big(F_1^{-1}(U), \ldots, F_d^{-1}(U)\big), \quad U \sim \mathcal{U}[0,1].$$

The symbol $\stackrel{d}{=}$ *means equality in distribution.*

Corollary 3 ($C = W_2 \Leftrightarrow$ countermonotonicity)

A bivariate random vector (X_1, X_2) *with marginals* F_1, F_2 *has copula* W_2 *if and only if*

$$(X_1, X_2) \stackrel{d}{=} \big(F_1^{-1}(U), F_2^{-1}(1-U)\big), \quad U \sim \mathcal{U}[0,1].$$

Basic rules for working with copulas

Computing probabilities from a distribution function

Given: The distribution function F of some random vector $(X_1, \ldots, X_d)$.

Aim: Calculate probabilities such as:

$$\mathbb{P}(a_1 < X_1 \leq b_1, \ldots, a_d < X_d \leq b_d), \quad -\infty < a_j < b_j < \infty,\ j = 1, \ldots, d.$$

Ansatz: The general formula is:

$$\begin{aligned}
\mathbb{P}(a_1 < X_1 \leq b_1, \ldots, a_d < X_d \leq b_d) &= \sum_{(c_1,\ldots,c_d) \in \times_{j=1}^{d} \{a_j, b_j\}} (-1)^{|\{j : c_j = a_j\}|} F(c_1, \ldots, c_d) \\
&= F(b_1, \ldots, b_d) - F(a_1, b_2, \ldots, b_d) - \ldots - F(b_1, \ldots, b_{d-1}, a_d) \\
&\quad + F(a_1, a_2, b_3, \ldots, b_d) + \ldots + F(b_1, \ldots, b_{d-2}, a_{d-1}, a_d) \\
&\quad - F(a_1, a_2, a_3, b_4, \ldots, b_d) - \ldots \ldots + (-1)^d F(a_1, \ldots, a_d).
\end{aligned}$$

Problem: Calculating this sum is numerically challenging.

Basic rules for working with copulas

Copula derivatives

Definition 7

A copula C is called "absolutely continuous", if it admits the integral representation

$$C(u_1,\dots,u_d)=\int_0^{u_1}\int_0^{u_2}\dots\int_0^{u_d} c(v_1,\dots,v_d)\,dv_d\,dv_{d-1}\dots dv_1,$$

for a non-negative function $c:(0,1)^d\to[0,\infty)$*, called the "(copula) density" of* C*.*

Remark 7

- *It follows that the density of C – provided it exists – can be computed as*

$$c(u_1,\dots,u_d)=\frac{\partial}{\partial u_1}\frac{\partial}{\partial u_2}\dots\frac{\partial}{\partial u_d}C(u_1,\dots,u_d). \tag{4}$$

- *Compared to the copula, the copula density has the advantage that it visualizes nicely where the probability mass is located.*

Basic rules for working with copulas

Copula derivatives

Example 2 (The bivariate Gaussian copula)

The most prominent absolutely continuous copula is the bivariate "Gaussian copula", which is defined in integral form by:

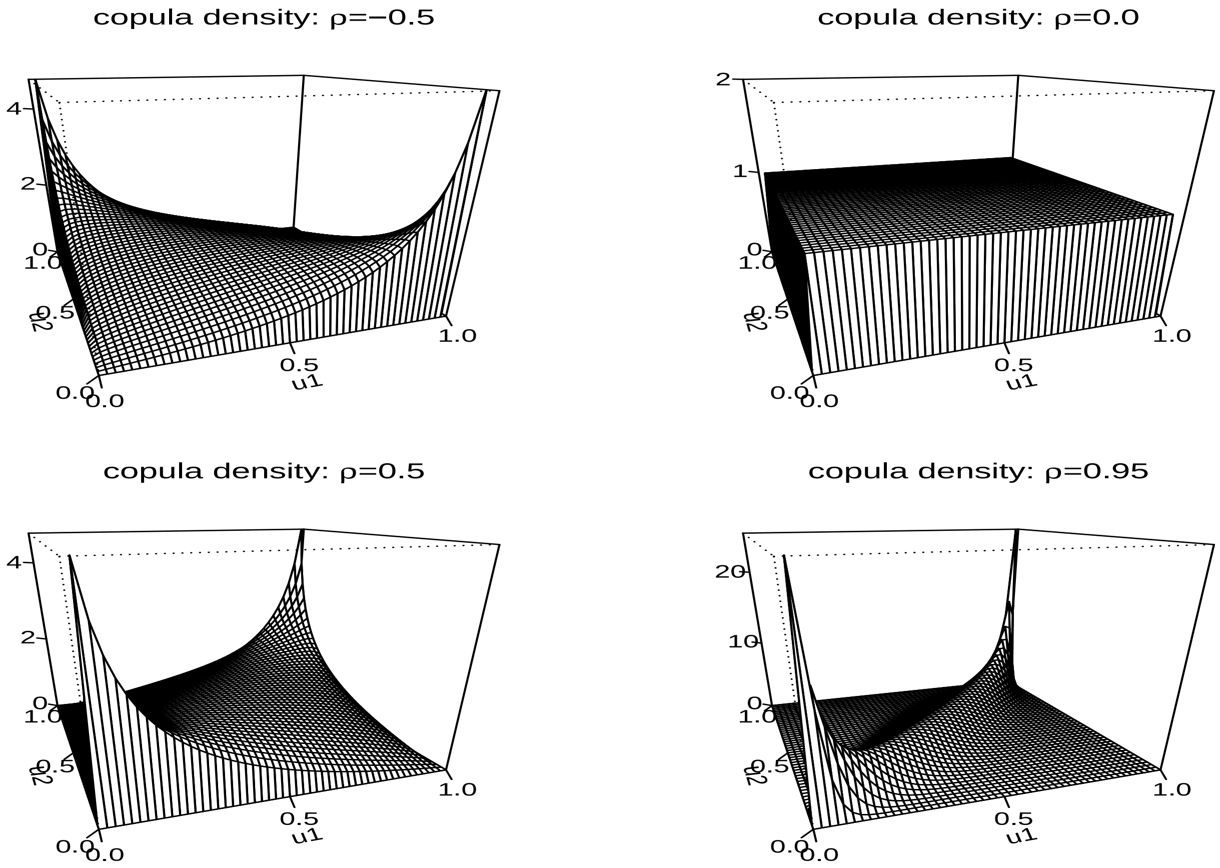
$$C_\rho(u_1, u_2) = \int_0^{u_1} \int_0^{u_2} c_\rho(v_1, v_2)\, dv_2\, dv_1,$$

$$c_\rho(u_1, u_2) = \frac{1}{\sqrt{1-\rho^2}} \exp\left(\frac{2\,\rho\,\Phi^{-1}(u_1)\,\Phi^{-1}(u_2) - \rho^2\left(\Phi^{-1}(u_1)^2 + \Phi^{-1}(u_2)^2\right)}{2\,(1-\rho^2)}\right),$$

for a dependence parameter $\rho \in (-1, 1)$. $\Phi(x) := \int_{-\infty}^{x} \exp(-y^2/2)\, dy / \sqrt{2\,\pi}$ *denotes the distribution function of a standard normally distributed random variable.*

Basic rules for working with copulas

Copula derivatives



Copula density $c_\rho(u_1, u_2)$ of a bivariate Gaussian copula for increasing ρ.

How to measure dependence?

Problem: Dependence is not a simple mathematical object, making it difficult to communicate information like the “degree-”, “level-”, or “type-of-dependence”.

Simplification: Compress information about the dependence structure into a single number that quantifies the degree of dependence on some scale (e.g. -1 to $+1$).

- Mapping from the set of copulas to the real numbers (one loses information).
- Several concepts exist, each covering only a certain aspect of the dependence structure (e.g. Pearson’s correlation: linear dependence). Which one to choose depends on the application.
- Dependence measures can be used to estimate parameters of copulas by comparing a theoretical dependence measure with the empirical counterpart.
- For several dependence measures we have empirical estimates with known finite sample (or asymptotic) distribution (useful, e.g. for hypothesis tests).

Pearson's correlation coefficient

Definition 8 (Pearson's correlation coefficient and its sample version)

Consider the random vector (X_1, X_2) and assume X_1 and X_2 are square integrable.

1. *"Pearson's correlation coefficient" is defined as*

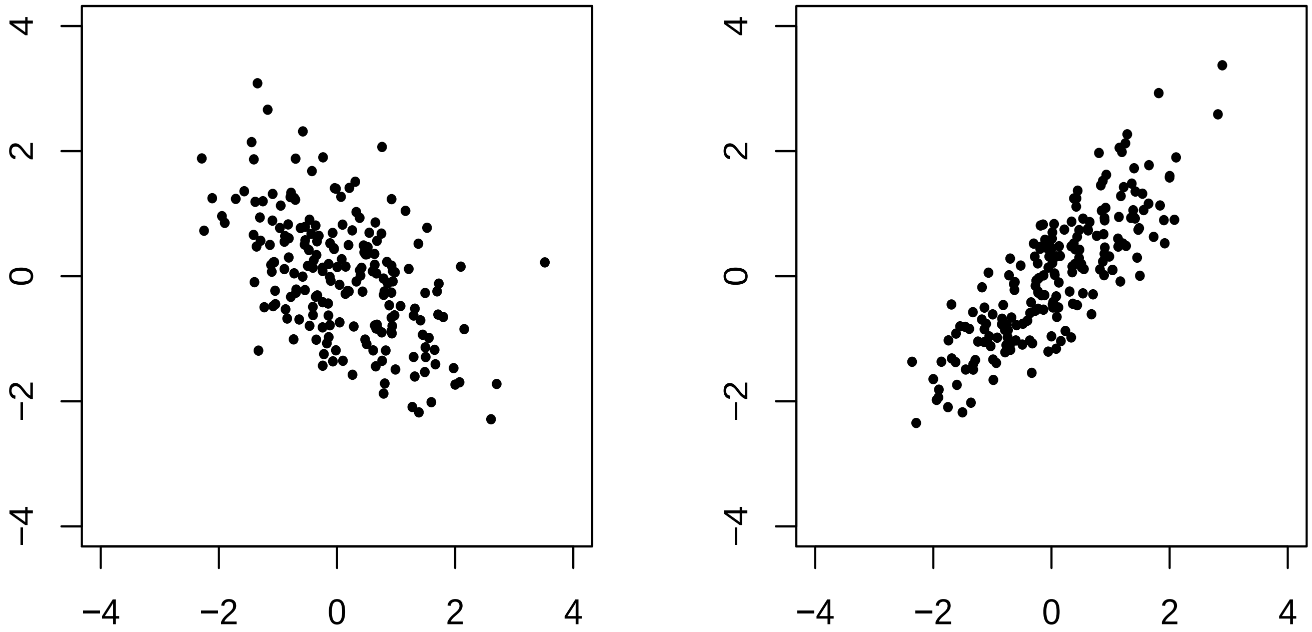
$$\rho = \mathrm{cor}(X_1, X_2) := \frac{\mathrm{cov}(X_1, X_2)}{\sqrt{\mathrm{Var}(X_1)}\sqrt{\mathrm{Var}(X_2)}} = \frac{\mathbb{E}[(X_1 - \mathbb{E}[X_1])(X_2 - \mathbb{E}[X_2])]}{\sqrt{\mathbb{E}[(X_1 - \mathbb{E}[X_1])^2]}\sqrt{\mathbb{E}[(X_2 - \mathbb{E}[X_2])^2]}}.$$

2. *Given iid observations $(X_1^{(1)}, X_2^{(1)}), \ldots, (X_1^{(n)}, X_2^{(n)})$ from (X_1, X_2), the empirical (or sample) estimate for the correlation is*

$$\hat{\rho}_n := \frac{\sum_{i=1}^n (X_1^{(i)} - \bar{X}_1)(X_2^{(i)} - \bar{X}_2)}{\sqrt{\sum_{i=1}^n (X_1^{(i)} - \bar{X}_1)^2}\sqrt{\sum_{i=1}^n (X_2^{(i)} - \bar{X}_2)^2}},$$

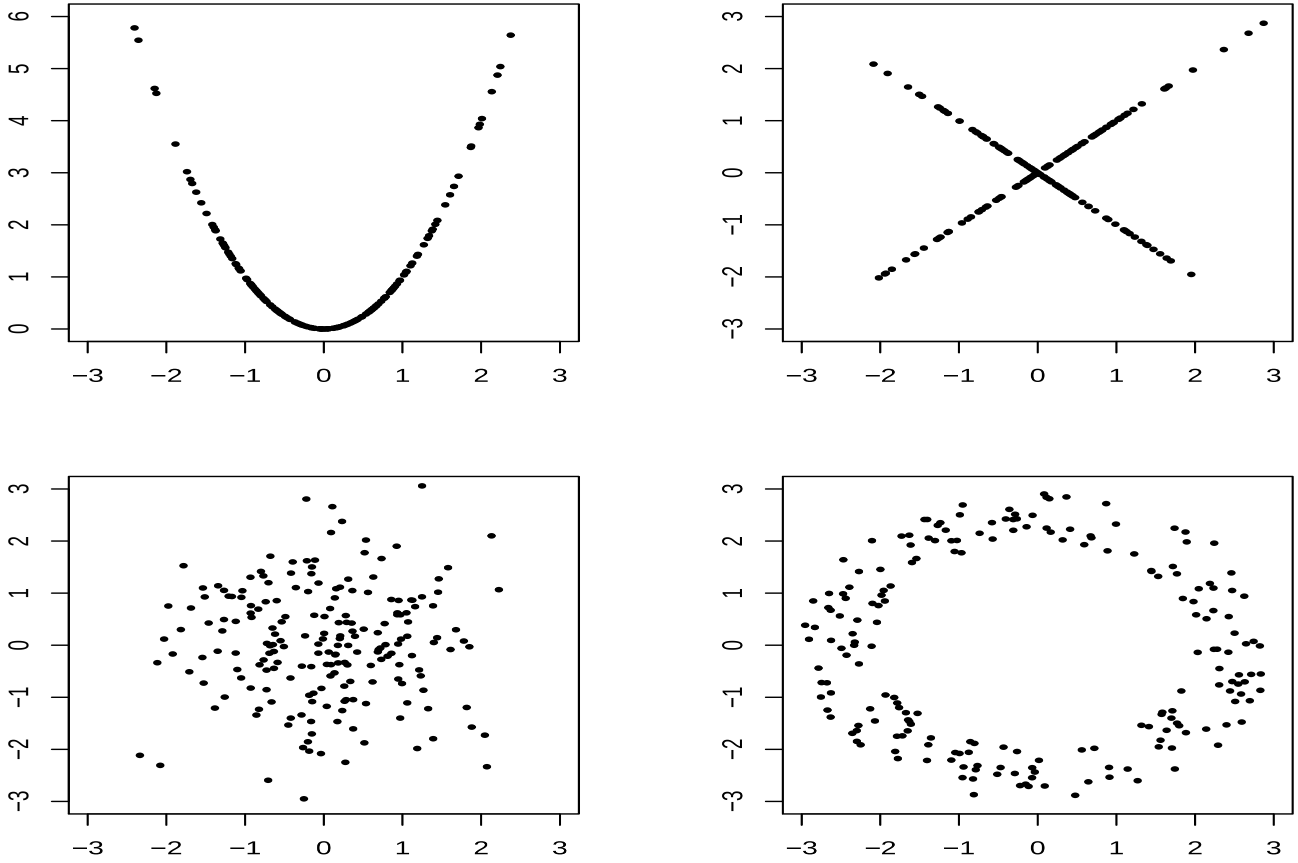
where $\bar{X}_j := \frac{1}{n}\sum_{i=1}^n X_j^{(i)}$, $j = 1, 2$.

Pearson’s correlation coefficient



$n = 200$ samples of a bivariate standard normal distribution with true correlation $\rho = -0.5$ (left) and $\rho = 0.8$ (right). The empirical estimates are around $\hat{\rho}_n \approx -0.52$ (left) and $\hat{\rho}_n \approx 0.83$ (right).

Pearson’s correlation coefficient



Scatter plots of four situations, where in each case the theoretical correlation is zero.

Concordance measures

Definition 9 (Concordant / discordant pairs)

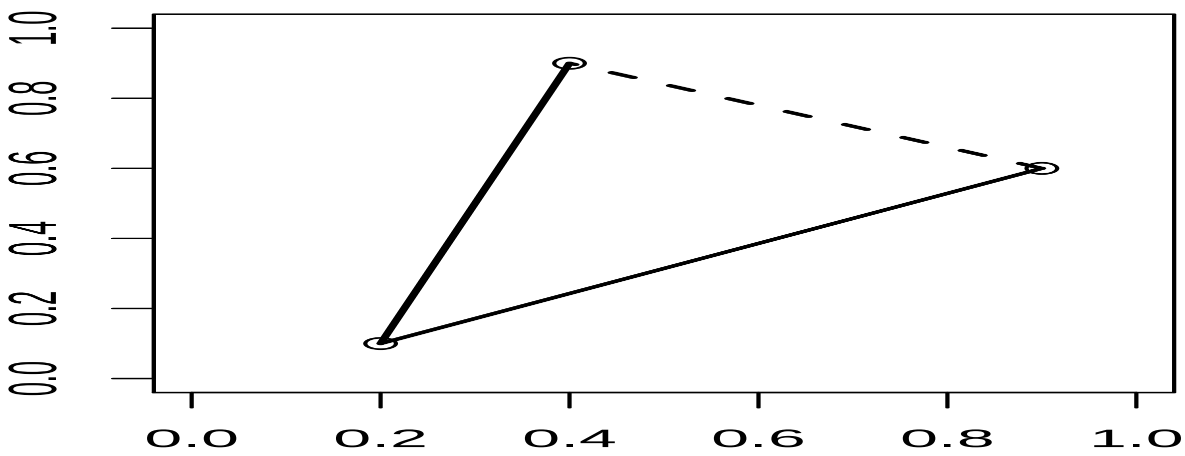
We say that (x_1, x_2) *and* (y_1, y_2) *are concordant if*

$$(x_1 - y_1)(x_2 - y_2) > 0$$

resp. discordant if $(x_1 - y_1)(x_2 - y_2) < 0$.

Example 3

To visualize concordance connect points with straight lines. Whenever the connecting line of a pair has positive slope we have concordance. Each concordant (discordant) pair is connected with a solid (dashed) line.



Concordance measures

Definition 10 (Kendall's τ)

1. *Consider the random vector* (U_1, U_2) *with copula* C *as joint distribution function. Then, "Kendall's* τ*" is defined as*

$$\tau = \tau_C := 4 \int_0^1 \int_0^1 C(u_1, u_2)\, dC(u_1, u_2) - 1 = 4\,\mathbb{E}\big[C(U_1, U_2)\big] - 1. \tag{5}$$

2. *For general bivariate random vectors* (X_1, X_2) *with continuous marginals, Kendall's* τ *is defined by applying the above Equation (5) to the unique copula of* (X_1, X_2)*, irrespectively of the marginals.*

Advantage: This is only a function of the copula, the marginals are not involved.

Question: Is there a link to concordance and discordance?

Concordance measures

Lemma 4 (Original definition, properties and empirical version of Kendall's τ)

(a) *Let $(U_1, U_2) \sim C$ and $(V_1, V_2) \sim C$ be independent. Then Kendall's τ equals*

$$\tau = \mathbb{P}\big(\underbrace{(U_1 - V_1)(U_2 - V_2) > 0}_{\text{concordance}}\big) - \mathbb{P}\big(\underbrace{(U_1 - V_1)(U_2 - V_2) < 0}_{\text{discordance}}\big).$$

– *Empirical version for iid samples $(X_1^{(1)}, X_2^{(1)}), \ldots, (X_1^{(n)}, X_2^{(n)})$:*

$$\begin{aligned}\hat{\tau}_n &:= \frac{\text{\# of concordant pairs} - \text{\# of discordant pairs}}{\text{\# of all pairs}} \\ &= \frac{\sum_{1 \le i < j \le n} \text{sign}\big[(X_1^{(j)} - X_1^{(i)})(X_2^{(j)} - X_2^{(i)})\big]}{n(n-1)/2}.\end{aligned}$$

– *For data with ties, there exist modified versions.*

– *For abs. continuous margins F_1 and F_2 we have $(X_1 - Y_1)(X_2 - Y_2) > 0$ if and only if $\big(F_1(X_1) - F_1(Y_1)\big)\big(F_2(X_2) - F_2(Y_2)\big) > 0$ (follows from F_1, F_2 strictly increasing), i.e. Kendall's τ only depends on C, not on F_1, F_2.*

Concordance measures

Lemma 4 (Original definition, properties and empirical version of Kendall's τ) (cont.)

(b) *Kendall's τ is increasing in the point-wise ordering of copulas:*

- *If $C(u_1, u_2) \leq \tilde{C}(u_1, u_2)$ for all $(u_1, u_2) \in [0, 1]^2$ then $\tau_C \leq \tau_{\tilde{C}}$.*

Moreover,

- *Kendall's τ of the independence copula is zero: $\tau_{\Pi_2} = 0$.*
- *$\tau_C = 1$ if and only if $C = M_2$ (comonotonicity copula).*

See [Nelsen (2006), Theorem 5.1.9] for a proof that Kendall's τ is a measure of concordance and hence satisfies these properties.

Concordance measures

Lemma 4 (Original definition, properties and empirical version of Kendall's τ) (cont.)

(c) *There exist reformulations of the analytical expression:*

$$\begin{aligned}
\tau_C &= 1 - 4\int_0^1\int_0^1 \frac{\partial}{\partial u_1}C(u_1,u_2)\,\frac{\partial}{\partial u_2}C(u_1,u_2)du_1\,du_2\\
&= 4\int_0^1\int_0^1 C(u_1,u_2)\,\frac{\partial^2}{\partial u_1\,\partial u_2}C(u_1,u_2)\,du_1\,du_2 - 1,
\end{aligned}$$

where the last expression requires C to be absolutely continuous.

Concordance measures

Another quite popular dependence measure is Speaman's ρ_S.

Definition 11 (Spearman's ρ_S)

Let (X_1, X_2) be a random vector with continuous marginal laws $X_j \sim F_j$. Define

$$(U_1, U_2) := \big(F_1(X_1), F_2(X_2)\big).$$

Then, "Spearman's ρ_S" is defined as Pearson's correlation coefficient of (U_1, U_2), i.e.

$$\rho_S := \rho_{S,C} = \operatorname{cor}(U_1, U_2) = \operatorname{cor}\big(F_1(X_1), F_2(X_2)\big). \tag{6}$$

Advantages:

- Spearman's ρ_S does not depend on the marginal laws F_j, $j = 1, 2$ (their influence is removed by the transformation to uniform marginals).
- Unlike for Pearson's correlation, we do not have to worry about existence of ρ_S, since $U_j \sim \mathcal{U}[0, 1]$, $j = 1, 2$ are square integrable.

Concordance measures

Lemma 5 (Properties of Spearman's ρ_S and its empirical version)

(a) – *Symmetry:* (X_1, X_2) *and* (X_2, X_1) *have the same* ρ_S.

– *Spearman's* ρ_S *is zero for the independence copula:* $\rho_{S,\Pi_2} = 0$.

– $\rho_{S,C} = 1$ *if and only if* $C = M_2$ *(comonotonicity copula).*

– *Again, we have ordering according to the point-wise ordering of copulas, i.e.* $C(u_1, u_2) \leq \tilde{C}(u_1, u_2)$ *for all* $(u_1, u_2) \in [0, 1]^2$ *implies*

$$\rho_{S,C} \leq \rho_{S,\tilde{C}}.$$

– *Let* C *be a copula and* $\hat{C}$ *its survival copula. Then*

$$\rho_{S,C} = \rho_{S,\hat{C}}.$$

See [Nelsen (2006), Theorem 5.1.9] for the proofs.

Concordance measures

Lemma 5 (Properties of Spearman's ρ_S and its empirical version) (cont.)

(b) *Consider an iid sample $\left(X_1^{(1)}, X_2^{(1)}\right), \ldots, \left(X_1^{(n)}, X_2^{(n)}\right)$ from (X_1, X_2) (with continuous margins, to avoid ties). The empirical (or sample) estimate of Spearman's ρ_S is the empirical correlation of the rank statistics of the sample values*

$$\hat{\rho}_{S,n} := \frac{\sum_{i=1}^{n} \left(\mathrm{rank}(X_1^{(i)}) - \frac{n+1}{2}\right) \left(\mathrm{rank}(X_2^{(i)}) - \frac{n+1}{2}\right)}{\sqrt{\sum_{i=1}^{n} \left(\mathrm{rank}(X_1^{(i)}) - \frac{n+1}{2}\right)^2} \sqrt{\sum_{i=1}^{n} \left(\mathrm{rank}(X_2^{(i)}) - \frac{n+1}{2}\right)^2}},$$

Again, this becomes more involved in the presence of ties.

(c) *Equivalent definitions:*

$$\begin{aligned}\rho_{S,C} &= 12 \int_0^1 \int_0^1 \left(C(u_1, u_2) - u_1\, u_2\right) du_1\, du_2 \\ &= 3\left(\mathbb{P}\left((U_1 - V_1)\,(U_2 - W_2) > 0\right) - \mathbb{P}\left((U_1 - V_1)\,(U_2 - W_2) < 0\right)\right),\end{aligned}$$

for independent copies (U_1, U_2), (V_1, V_2), and (W_1, W_2) with distribution fct. C.

Concordance measures

Remark 8

The term $(n+1)/2$ *in the definition of* $\hat{\rho}_{S,n}$ *is simply the averages of the rank statistics – if you add up the first n ranks, this is the same as adding up the natural numbers until n. Deviding by n to get the average yields the result.*

Example 4 ("Ranks")

Consider the observations in the Table. The empirical Spearman's ρ_S *is* $\hat{\rho}_{S,5} = 0.7$*.*

i	1	2	3	4	5	i	1	2	3	4	5
$X_1^{(i)}$	1.1	2.3	4.9	0.5	5.5	$X_2^{(i)}$	0.9	1.2	5.2	3.3	6.0
$\mathrm{rank}(X_1^{(i)})$	2	3	4	1	5	$\mathrm{rank}(X_2^{(i)})$	1	2	4	3	5

Concordance measures

Applications of Kendall's τ and Spearman's ρ_S:

(a) **Dependence measuring:** Measure the strength of dependence implied by some copula or being empirically observed in some set of data.

(b) **Testing for independence:** Use the empirical versions of Kendall's τ and Spearman's ρ_S to test the hypothesis $\mathcal{H}_0$: X_1 and X_2 are independent.

(c) **Parameter estimation:** For a bivariate copula from a one-parameter family, express Kendall's τ and Spearman's ρ_S as functions of the parameter and estimate the parameter using the empirical Kendall's τ and Spearman's ρ_S.

Concordance measures

(b) Testing for independence:

- **Given**: iid observations $\left(X_1^{(1)}, X_2^{(1)}\right), \ldots, \left(X_1^{(n)}, X_2^{(n)}\right)$ from (X_1, X_2).
- **Test hypothesis**: $\mathcal{H}_0$: X_1 and X_2 are independent.
- **Approach**: Test if $\hat{\tau}_n$ and $\hat{\rho}_{S,n}$ are significantly different from zero. In that case we reject $\mathcal{H}_0$.
- **Intuition**: If $\mathcal{H}_0$ is correct, then the empirical versions $\hat{\tau}_n$ and $\hat{\rho}_{S,n}$ must be “close to” zero, since this is the theoretical value under $\mathcal{H}_0$.
  - Exact (or asymptotic for big n) distribution of $\hat{\tau}_n$ and $\hat{\rho}_{S,n}$ is needed.

Concordance measures

(c) Parameter estimation:

- **Situation**: Many families of bivariate copulas are parameterized by a single parameter, θ. Let the copula of (X_1, X_2) be from such a one-parameter family.
- **Aim**: Estimate θ given iid observations $\left(X_1^{(1)}, X_2^{(1)}\right), \dots, \left(X_1^{(n)}, X_2^{(n)}\right)$ from (X_1, X_2).
- **Approach**:

(i) Express Kendall's τ and Spearman's ρ_S as functions of this parameter, $\tau = f(\theta)$ and $\rho_S = g(\theta)$. In most cases f and g have inverses f^{-1} and g^{-1}.

(ii) Calculate the empirical versions of Kendall's τ or Spearman's ρ_S, $\hat{\tau}_n$ or $\hat{\rho}_{S,n}$ from the sample.

(iii) Use $\hat{\theta}_n := f^{-1}(\hat{\tau}_n)$ or $\hat{\theta}_n := g^{-1}(\hat{\rho}_{S,n})$ as an estimator for θ. This estimation methodology will later be explained in more detail.

Concordance measures

Motivation of tail dependence:

- In financial applications we are often concerned with problems such as:
  - Modeling the probability of joint defaults in credit portfolios where each default event has a small probability.
  - Joint drop of two (or more) stocks.
- In both cases, not the “center of the joint distribution” matters, but the “tails”.
- Often, it is reasonable to argue that dependence increases for adverse events.
  - Possible reasons: herd behavior (panic selling), technical (broken limits), ...
  - Thus, diversification often breaks down just when it is needed the most.
- Tail dependence relates to questions like “**given X_1 is extreme, what is the conditional probability of X_2 being also extreme?**”.

Tail dependence

Definition 12 (Tail dependence)

The lower- and upper-tail dependence coefficients of (X_1, X_2) *with copula* C *are*

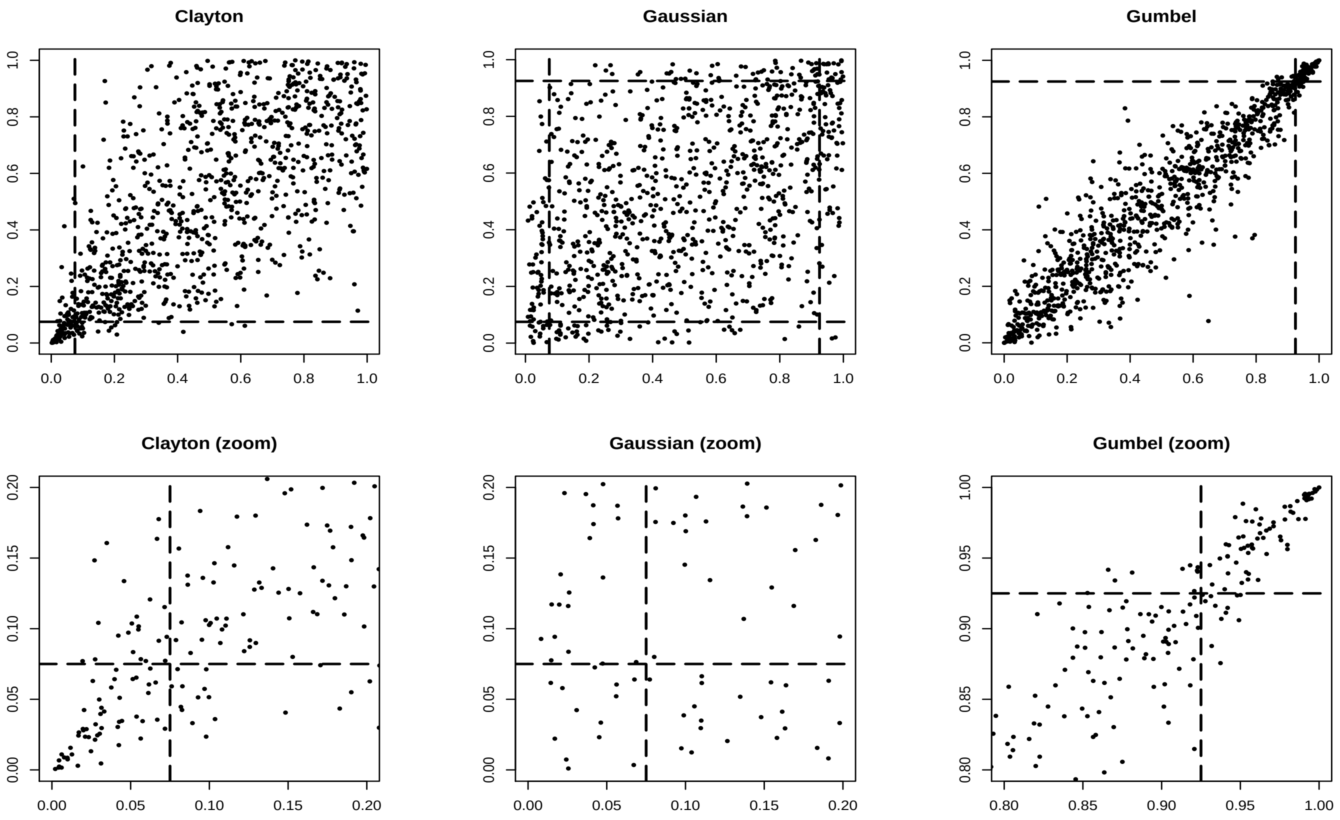
$$LTD_C := \lim_{\alpha \searrow 0} \mathbb{P}\big(X_1 \leq F_1^{-1}(\alpha) | X_2 \leq F_2^{-1}(\alpha)\big) = \lim_{u \searrow 0} \frac{C(u,u)}{u}, \tag{7}$$

$$UTD_C := \lim_{\alpha \nearrow 1} \mathbb{P}\big(X_1 > F_1^{-1}(\alpha) | X_2 > F_2^{-1}(\alpha)\big) = \lim_{u \nearrow 1} \frac{C(u,u) - 2u + 1}{1-u}, \tag{8}$$

provided that these limits exist.

Dependence for “black-swan events”.

Tail dependence



Scatter plots of three different copulas and zoom into the corners. Clayton: positive LTD, zero UTD; Gaussian: zero LTD, zero UTD; Gumbel: zero LTD, positive UTD.

Lecture II

Archimedean and elliptical copulas

Archimedean copulas

Definition 13 (Archimedean copula)

A copula $C_\varphi : [0,1]^d \to [0,1]$ is an ***Archimedean copula*** *if it has the functional form*

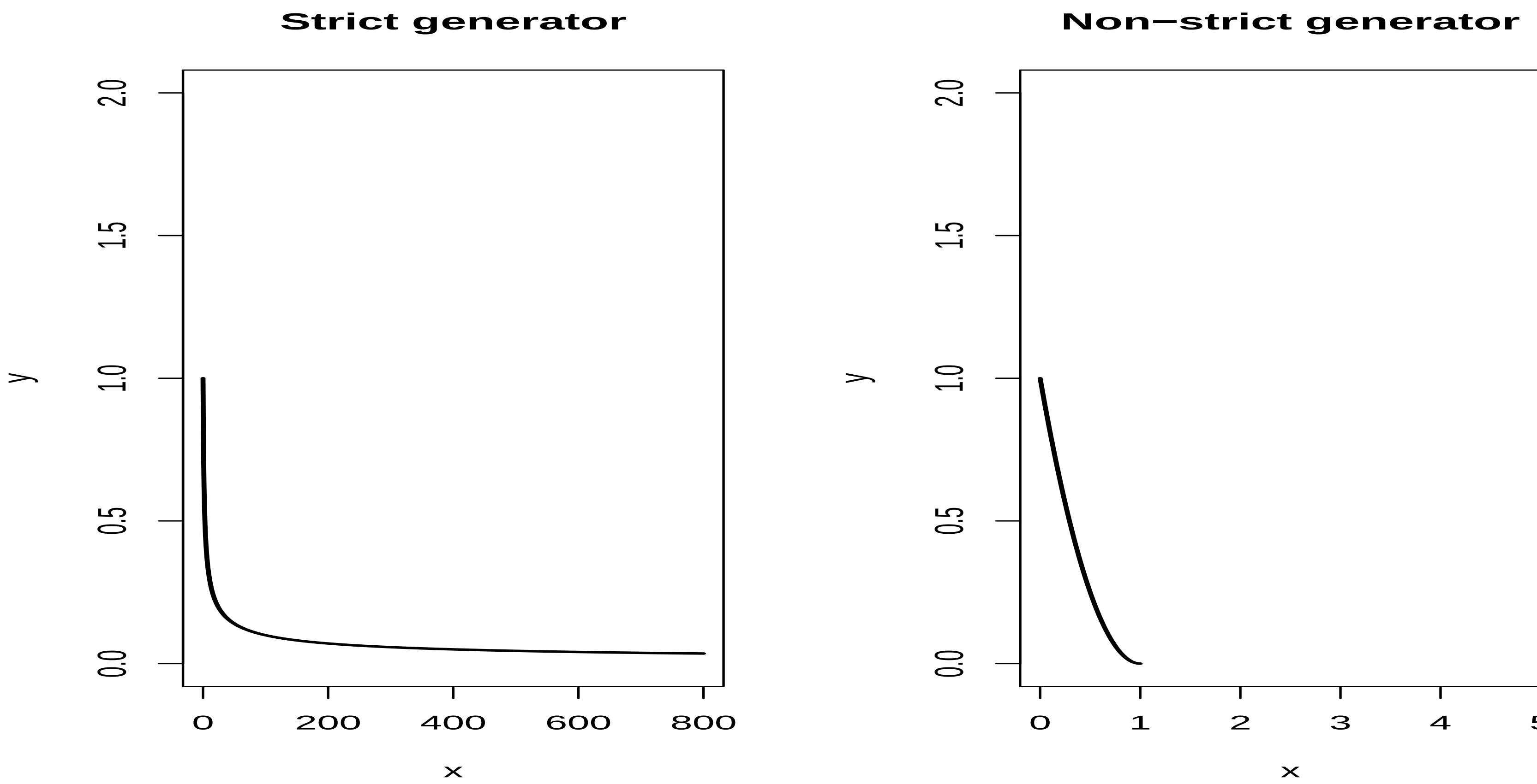
$$C_\varphi(u_1, \ldots, u_d) = \varphi\big(\varphi^{-1}(u_1) + \ldots + \varphi^{-1}(u_d)\big), \tag{9}$$

*for a suitable, non-increasing function $\varphi : [0, \infty) \to [0,1]$ with $\varphi(0) = 1$ and $\lim_{x\to\infty} \varphi(x) = 0$, called "**(Archimedean) generator**".*

Example 5 (The independence copula is an Archimedean copula)

The function $\varphi(x) = \exp(-x)$ is an Archimedean generator, $\varphi^{-1}(y) = -\log(y)$. Plugging it into Equation (9), we observe that $C_\varphi = \Pi$.

Archimedean copulas: Generator



Archimedean copulas: Why “Archimedean“ copulas?

Archimedean Axiom

$$\forall a, b \in \mathbb{R}^+ \text{ such that } a < b \; \exists n \in \mathbb{N} \text{ with } na > b$$

Define C-powers u_C^n of u recursively:

$$\begin{aligned} u_C^1 &= u \\ u_C^{n+1} &= C(u, u_C^n) \end{aligned}$$

Archimedean Axiom for copulas

Let C be an Archimedean copula generated by φ.

Then for any $u, v \in (0,1)$ such that $u > v$ $\exists n \in \mathbb{N}$ with $u_C^n < v$.

Proof.

By induction is $u_C^n = \varphi\left[n\varphi^{-1}(u)\right]$.

Since $\varphi^{-1}(u), \varphi^{-1}(v) > 0 \;\Rightarrow\; \exists n \in \mathbb{N}$ such that $n\varphi^{-1}(u) > \varphi^{-1}(v)$.

But since $v > 0$,$\varphi(v) < \varphi(0)$, and hence:

$$v = \varphi\left[\varphi^{-1}(v)\right] > \varphi\left[n\varphi^{-1}(u)\right] = u_C^n$$

Archimedean copulas

Definition 14 (Completely monotone generator)

Φ_∞ denotes the set of all "completely monotone" generators, i.e. all φ with:

- *φ is continuous at zero and $\varphi(0) = 1$,*
- *φ is infinitely often differentiable on the interior of its domain $(0, \infty)$, and*
- *the derivatives satisfy $(-1)^k \varphi^{(k)}(x) \geq 0$ for all $x > 0$, $k \in \mathbb{N}_0$, where $\varphi^{(k)}$ denotes the k-th derivative of φ and $\varphi^{(0)} := \varphi$.*

Lemma 6

Let $\varphi \in \Phi_\infty$. Then C_φ is a proper distribution function in each dimension $d \geq 2$.

Archimedean copulas

Remark 9 (Parameterization of Archimedean copulas)

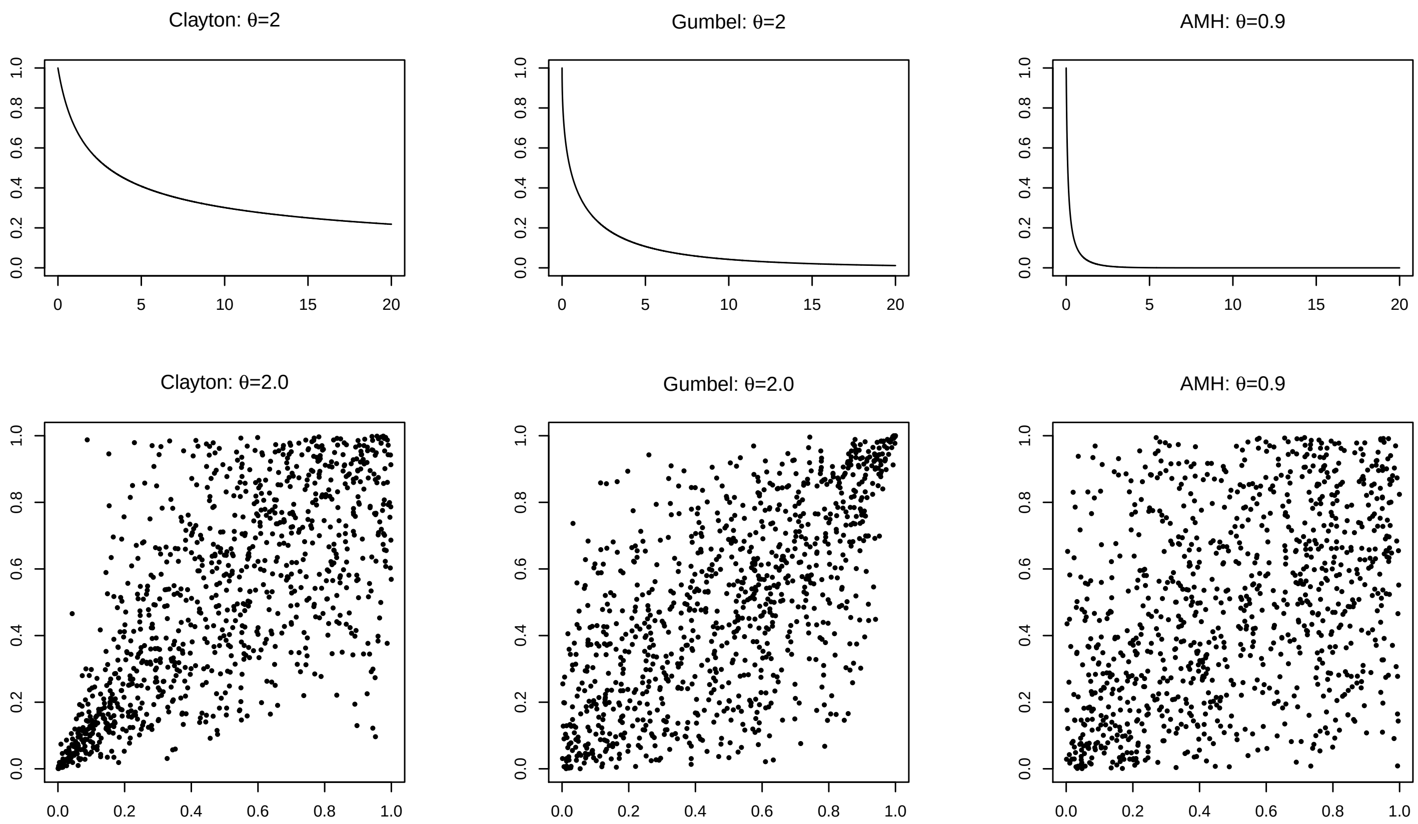
- *Archimedean copulas form an infinite-dimensional space, because they are parameterized by a function φ instead of parameters.*
- *However, in practice, one usually chooses a parametric family of Laplace transforms, i.e. $\varphi = \varphi_\theta$ for a real parameter θ. In this case, we write $C_{\varphi_\theta} = C_\theta$.*

Example 6 (Some popular Archimedean copulas)

The table gathers the most popular Archimedean copulas and their generators.

$\varphi_\theta(x)$	$\varphi_\theta^{-1}(y)$	$\theta \in$	*name of copula*	*Kendall's τ*
$(1+x)^{-1/\theta}$	$y^{-\theta} - 1$	$(0, \infty)$	*Clayton*	$\theta/(2+\theta)$
$e^{-x^{1/\theta}}$	$\big(-\log(y)\big)^\theta$	$[1, \infty)$	*Gumbel*	$(\theta-1)/\theta$
$\frac{1-\theta}{e^x-\theta}$	$\log\left(\frac{1-\theta}{y}+\theta\right)$	$[0, 1)$	*Ali–Mikhail–Haq*	$1-2\left(\theta+(1-\theta)^2\log(1-\theta)\right)/(3\,\theta^2)$

Archimedean copulas



Generators (top) and scatter plots (bottom) for the Clayton, Gumbel, and Ali–Mikhail–Haq (AMH) copula.

Archimedean copulas

Remark 10 (Important stylized facts of Archimedean copulas)

(a) ***Dependence range***:

- *One can show that for every Laplace transform φ, it holds true that $C_\varphi \geq \Pi$ pointwise:*
  - *Negative dependence cannot be modeled by Archimedean copulas with completely monotone generators.*
  - *Concordance measures are non-negative.*
- *Typical Archimedean families include the independence copula Π and the upper Fréchet–Hoeffding bound as boundary cases.*

(b) ***Symmetries***:

- *Archimedean copulas are exchangeable, due to their algebraic expression, even for $d > 2$.*

Archimedean copulas

Remark 10 (Important stylized facts of Archimedean copulas) (cont.)

(c) **Concordance measures**:

- Formulas for concordance measures of arbitrary Archimedean copulas are only given in terms of an integral involving the function φ:

$$\tau = 1 - 4 \int_0^\infty x \cdot \left(\varphi^{(1)}(x)\right)^2 dx, \qquad \text{(Kendall's } \tau\text{)} \tag{10}$$

$$\rho_S = 12 \int_0^1 \int_0^1 C_\varphi(u_1, u_2)\, du_2\, du_1 - 3. \qquad \text{(Spearman's } \rho_S\text{)} \tag{11}$$

The formula for Spearman's ρ_S is actually valid for any copula, not only Archimedean ones, see Definition 11.

- Whether this can be computed in closed form depends on the generator.

Archimedean copulas

Remark 10 (Important stylized facts of Archimedean copulas) (cont.)

(d) **Tail dependence coefficients**:

- The upper- and lower-tail dependence coefficients are given by the following formulas – provided the respective limits exist:

$$LTD_{C_\varphi} = 2 \cdot \lim_{x \nearrow \infty} \frac{\varphi^{(1)}(2\,x)}{\varphi^{(1)}(x)} = \lim_{x \nearrow \infty} \frac{\varphi(2\,x)}{\varphi(x)},$$
$$UTD_{C_\varphi} = 2 - 2 \cdot \lim_{x \downarrow 0} \frac{\varphi^{(1)}(2\,x)}{\varphi^{(1)}(x)}.$$

  – Archimedean copulas allow for asymmetric tail dependence coefficients.

  – Revisiting the examples from the table:

$$\begin{aligned} \text{Clayton:} \quad & LTD_\theta = 2^{-1/\theta}, \quad UTD_{C_\theta} = 0, \\ \text{Gumbel:} \quad & LTD_\theta = 0, \quad UTD_\theta = 2 - 2^{1/\theta}, \\ \text{Ali–Mikhail–Haq:} \quad & LTD_\theta = 0, \quad UTD_\theta = 0, \end{aligned}$$

Archimedean copulas

Remark 10 (Important stylized facts of Archimedean copulas) (cont.)

(e) **Density**:

- Archimedean copulas with completely monotone generator are absolutely continuous.
  - The density is obtained by taking iteratively the partial derivatives of $C_\varphi(u_1, \ldots, u_d)$ with respect to all components $u_1, \ldots, u_d$.
  - In dimension $d = 2$, this yields the density:

$$c_\varphi(u_1, u_2) = \frac{\partial^2}{\partial u_1 \, \partial u_2} C_\varphi(u_1, u_2) = \frac{\varphi^{(2)}\left(\varphi^{-1}(u_1) + \varphi^{-1}(u_2)\right)}{\varphi^{(1)}\left(\varphi^{-1}(u_1)\right) \varphi^{(1)}\left(\varphi^{-1}(u_2)\right)}, \quad u_1, u_2 \in (0, 1).$$

- Computing the density in larger dimensions $d \geq 2$ becomes burdensome due to the involved d-fold derivative $\varphi^{(d)}$.

Elliptical copulas: Elliptical distributions

Definition 7

Let $\mathbb{S}_d$ denote the space of all symmetric $d \times d$ matrices. A random vector $\mathbf{X} = (X_1, \ldots, X_d)' \in \mathbb{R}^d$ is said to have an (non-degenerate) elliptical distribution with parameters $\mu \in \mathbb{R}^d$ and $\Sigma = (\sigma_{k\ell})_{k,\ell \in \{1,\ldots,d\}} \in \mathbb{S}_d$, if

$$\mathbf{X} = \mu + A\mathbf{Y},$$

where $\mathbf{Y}$ has a m-dimensional spherical distribution and A is $d \times m$ matrix such that $AA' = \Sigma$ with $\mathrm{rank}(\Sigma) = m$.

Remark 11

A random vector $\mathbf{X} = (X_1, \ldots, X_d)' \in \mathbb{R}^d$ is said to have an (non-degenerate) elliptical distribution with parameters $\mu \in \mathbb{R}^d, \Sigma = (\sigma_{k\ell})_{k,\ell \in \{1,\ldots,d\}} \in \mathbb{S}_d$ and the generator function g, if its characteristic function $E(exp(i\mathbf{t}^\top \mathbf{X}))$ with $\mathbf{t} \in \mathbb{R}^d$ has the representation

$$exp(i\mathbf{t}^T \mu) g(t^T \Sigma \mathbf{t})$$

for some scalar function g.

Definition 8

Elliptical copulas are the copulas of elliptical distributions.

Gaussian copulas

Recall: If $Y_1, \ldots, Y_d$ are iid standard normally distributed random variables, $\mu_1, \ldots, \mu_d \in \mathbb{R}$, and $A = (a_{i,j}) \in \mathbb{R}^{d\times d}$ a matrix with full rank, the random vector

$$\mathbf{X} := \begin{pmatrix} X_1 \\ \vdots \\ X_d \end{pmatrix} = \begin{pmatrix} \mu_1 \\ \vdots \\ \mu_d \end{pmatrix} + A \cdot \begin{pmatrix} Y_1 \\ \vdots \\ Y_d \end{pmatrix} = \begin{pmatrix} \mu_1 + a_{1,1}\, Y_1 + \ldots + a_{1,d}\, Y_d \\ \vdots \\ \mu_d + a_{d,1}\, Y_1 + \ldots + a_{d,d}\, Y_d \end{pmatrix} \in \mathbb{R}^d \tag{12}$$

is said to have a **multivariate normal distribution**.

- **Marginals:** For each $j = 1, \ldots, d$, $X_j \sim \mathcal{N}(\mu_j, \sigma_j^2)$ with $\sigma_j^2 := \sum_{l=1}^d a_{j,l}^2$.
- **Correlation matrix:** Denote by $\Sigma := (\rho_{j,k})_{j,k=1,\ldots,d}$ the correlation matrix of $(X_1, \ldots, X_d)$, i.e. for $j, k = 1, \ldots, d$:
$$\rho_{j,k} = \mathrm{cor}(X_j, X_k).$$
- **Copula:** Since the marginal laws are continuous (univariate normals), the copula of $(X_1, \ldots, X_d)$ is unique by virtue of Sklar's Theorem 2.

Gaussian copulas

Definition 15 (Gaussian copula)

*The copulas of multivariate normal distributions, see the stochastic model from Equation (12), are called "**Gaussian copulas**".*

Remark 12 (The parameters of a Gaussian copula)

- *A Gaussian copula is independent of* $\mu_1, \ldots, \mu_d$ *and* $\sigma_1^2, \ldots, \sigma_d^2$.
- *Consequently, it is parameterized solely by* Σ *and we denote it by* C_Σ.
- *Thus the bivariate pairs / pair correlations already specify the copula.*

Gaussian copulas

Example 7 (The bivariate case)

For $d = 2$ the Gaussian copula depends on a single parameter $\rho := \rho_{1,2} = \rho_{2,1}$, due to symmetry of the correlation coefficient, since in this case

$$\Sigma = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}$$

Thus, we denote the bivariate Gaussian copula by C_ρ instead of C_Σ. It is given by

$$C_\rho(u_1, u_2) = \int_0^{u_1} \int_0^{u_2} \frac{\exp\left(\frac{2\,\rho\,\Phi^{-1}(v_1)\,\Phi^{-1}(v_2) - \rho^2\left(\Phi^{-1}(v_1)^2 + \Phi^{-1}(v_2)^2\right)}{2\,(1-\rho^2)}\right)}{\sqrt{1-\rho^2}}\, dv_2\, dv_1, \qquad (13)$$

where Φ denotes the standard normal distribution function.

***Observation**:*

(i) *C_ρ is absolutely continuous, the density is the integrand in Equation (13).*

(ii) *Both the numerical evaluation and the analytical study of the Gaussian copula are burdensome because of the appearing double integral.*

Gaussian copulas

The normal law is omnipresent in applications. Why so?

Some reasons are:

(a) **Natural appearance**:

- Consider a random vector $\mathbf{X}$ with existing mean vector μ and existing covariance matrix Σ.
- Let $\mathbf{X}^{(1)}, \ldots, \mathbf{X}^{(n)}$ be n iid samples from $\mathbf{X}$, e.g. the same experiment repeated n times.
- The (multivariate) central limit theorem states that the $\sqrt{n}$-scaled deviation from the mean

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} (\mathbf{X}^{(i)} - \mu)$$

has approximately a multivariate normal law with zero mean vector and covariance matrix Σ.

Gaussian copulas

The normal law is omnipresent in applications. Why so?

Some reasons are:

(b) **Mathematical tractability**:

- The multivariate normal distribution has an intrinsic, close connection to the theory of linear algebra.
- For instance, if $\mathbf{X}$ is multivariate normal with mean vector $\mu \in \mathbb{R}^d$ and covariance matrix $\Sigma \in \mathbb{R}^{d\times d}$, and $A \in \mathbb{R}^{m\times d}$, then $A\mathbf{X}$ is multivariate normal with mean vector $A\mu \in \mathbb{R}^m$ and covariance matrix $A\Sigma A' \in \mathbb{R}^{m\times m}$.
- Therefore, many applications can be deduced by resorting to the well-established apparatus of linear algebra.

Gaussian copulas

The normal law is omnipresent in applications. Why so?

Some reasons are:

(c) **Convenient parameterization**:

- The mean vector and covariance matrix specify the distribution completely.
- A finite number of parameters is a very convenient assumption for applications, in particular in large dimensions.
- It is not too difficult to construct low-parametric families of multivariate normal distributions even for very large dimensions.

Gaussian copulas

The normal law is omnipresent in applications. Why so?

Some reasons are:

(d) **Intuitive stochastic model**:

- The covariance matrix Σ specifies a certain “dispersion area” around the expected mean μ.
- Many applications rely on the idea of modeling an expected outcome and a dispersion around it, for which the normal distribution is a natural candidate.
- **Warning!** If the phenomenon to be modeled does not fit the interpretation of a dispersion around a mean, the use of a multivariate normal distribution model is actually not justified. However, in the past it has often been applied in such situations, especially in Finance.

Gaussian copulas

The normal law is omnipresent in applications. Why so?

Some reasons are:

(e) **Common ground**:

- Everyone knows the multivariate normal distribution.

As a consequence of these reasons, the multivariate normal distribution is by far the most popular distribution in financial (and many other) applications.

Gaussian copulas

Remark 13 (Important stylized facts of the bivariate Gaussian copula)

(a) ***Dependence range:***

- *With ρ ranging in $[-1, 1]$, the Gaussian copula C_ρ interpolates between the lower Fréchet–Hoeffding bound and the upper Fréchet–Hoeffding bound*

$$C_{-1} = W_2, \; C_0 = \Pi_2, \; \textit{and } C_1 = M_2.$$

- *This interpolation property allows to model the full spectrum of dependence and is a very desirable feature of the model.*
- *In particular, it provides the parameter ρ with an intuitive meaning: dependence increases with ρ.*

Gaussian copulas

Remark 13 (Important stylized facts of the bivariate Gaussian copula) (cont.)

(b) ***Concordance measures:*** *The following formulas are known for the bivariate Gaussian copula:*

$$\tau_\rho = \frac{2}{\pi}\,\text{arcsin}(\rho), \qquad \text{(Kendall's } \tau\text{)}$$

$$\rho_S = \frac{6}{\pi}\,\text{arcsin}(\rho/2), \qquad \text{(Spearman's } \rho_S\text{)}$$

$$\beta_\rho = \tau_\rho = \frac{2}{\pi}\,\text{arcsin}(\rho). \qquad \text{(Blomqvist's } \beta\text{)}$$

Gaussian copulas

Remark 13 (Important stylized facts of the bivariate Gaussian copula) (cont.)

(c) ***Symmetries:*** *The Gaussian copula exhibits two very strong symmetries.*

(i) *It is radially symmetric, i.e.* $C_\rho = \hat{C}_\rho$ *(even in all dimensions* d*).*

(ii) *The bivariate Gaussian copula is exchangeable, i.e.* $C_\rho(u_1, u_2) = C_\rho(u_2, u_1)$.

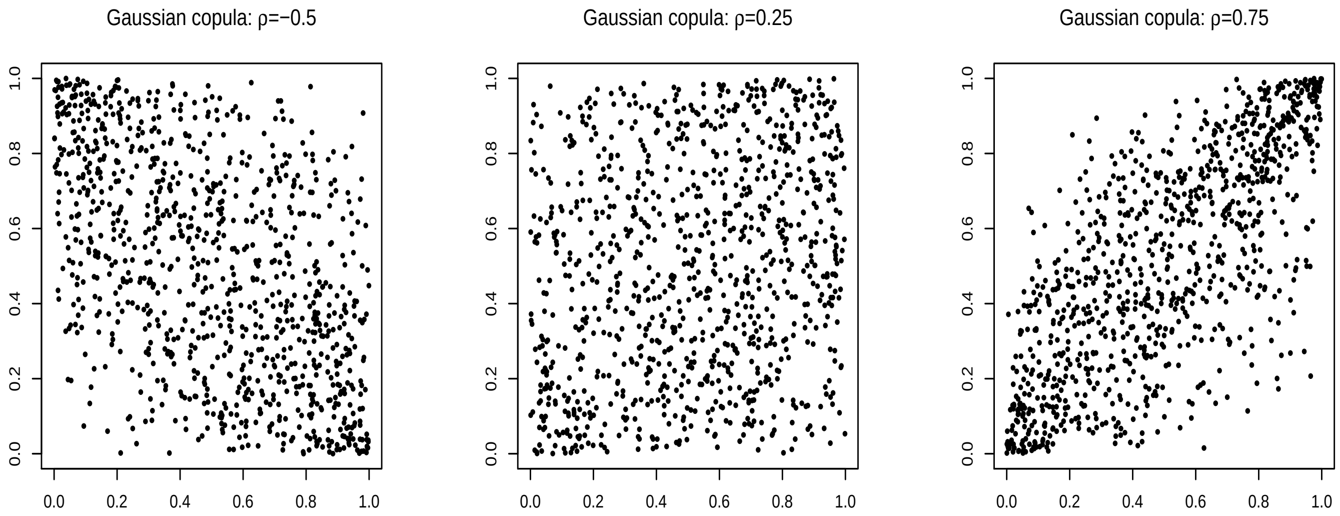
- *On scatter plots the points are scattered symmetrically around the diagonal* $\{(u_1, u_2) \in [0, 1]^2 \, : \, u_2 = u_1\}$.

In financial modeling, both symmetry properties can lead to serious problems.

Gaussian copulas

Remark 13 (Important stylized facts of the bivariate Gaussian copula) (cont.)

(c) ***Symmetries:*** *Scatter plots of the bivariate Gaussian copula with correlations* $\rho = -0.5$ *(left),* $\rho = 0.25$ *(middle), and* $\rho = 0.75$ *(right). Observe the symmetries.*



Gaussian copulas

Remark 13 (Important stylized facts of the bivariate Gaussian copula) (cont.)

(d) ***Tail independence:***

- *For every $\rho \in (-1, 1)$, the Gaussian copula exhibits tail independence, i.e. both, the upper- and the lower-tail dependence coefficient of the bivariate Gaussian copula are zero.*
- *This might not be desirable in the context of financial modeling.*

t-copulas

Definition 16 (Multivariate t-distribution)

Let $Y_1, \ldots, Y_d$ be iid standard normal random variables, and let $W \sim \mathrm{Inv}\Gamma(\nu/2, \nu/2)$ for some $\nu > 0$ be independent of $Y_1, \ldots, Y_d$. Moreover, let $\mu_1, \ldots, \mu_d \in \mathbb{R}$, and $A = (a_{i,j}) \in \mathbb{R}^{d \times d}$ with full rank. The random vector

$$\begin{pmatrix} X_1 \\ \vdots \\ X_d \end{pmatrix} = \begin{pmatrix} \mu_1 \\ \vdots \\ \mu_d \end{pmatrix} + A \cdot \sqrt{W} \begin{pmatrix} Y_1 \\ \vdots \\ Y_d \end{pmatrix} = \begin{pmatrix} \mu_1 + a_{1,1}\sqrt{W}\, Y_1 + \ldots + a_{1,d}\sqrt{W}\, Y_d \\ \vdots \\ \mu_d + a_{d,1}\sqrt{W}\, Y_1 + \ldots + a_{d,d}\sqrt{W}\, Y_d \end{pmatrix} \in \mathbb{R}^d \tag{14}$$

*is said to have a "**multivariate t-distribution**" with ν degrees of freedom.*

t-copulas

Remark 14 (Inverse Gamma distribution)

- *Comparing the stochastic models (14) with (12), the sole difference is the appearance of the inverse Gamma random variable W.*
- *A random variable W has an inverse Gamma distribution with parameters $\beta, \eta > 0$, we write $W \sim \mathrm{Inv}\Gamma(\beta, \eta)$, if W has probability density function*

$$f_W(x) = \mathbf{1}_{\{x>0\}} \frac{\eta^\beta \, e^{-\eta/x}}{x^{\beta+1} \, \Gamma(\beta)}.$$

t-copulas

- Like Gaussian copulas are derived from multivariate normal laws, t-copulas are associated with multivariate t-distributions via Sklar's Theorem.

- The t-copula only depends on the degrees of freedom ν and a correlation matrix $\Sigma \in \mathbb{R}^{d\times d}$, which is defined by

$$\Sigma_{i,j} := \frac{\sum_{k=1}^{d} a_{i,k}\, a_{j,k}}{\sqrt{\sum_{k=1}^{d} a_{i,k}^2 \sum_{k=1}^{d} a_{j,k}^2}}, \quad i,j = 1,\dots,d.$$

  We denote it by $C_{\nu,\Sigma}$.

  – $C_{\nu,\Sigma}$ is independent of the means $\mu_1,\dots,\mu_d$.
  – Be aware that Σ is not the correlation matrix of $(X_1,\dots,X_d)$.
  – The degrees of freedom ν also affect the final correlation matrix of

$$(X_1,\dots,X_d).$$

t-copulas

Example 8 (The bivariate case)

For $d = 2$, Σ has a single parameter ρ, and we denote the bivariate t-copula by $C_{\nu,\rho}$:

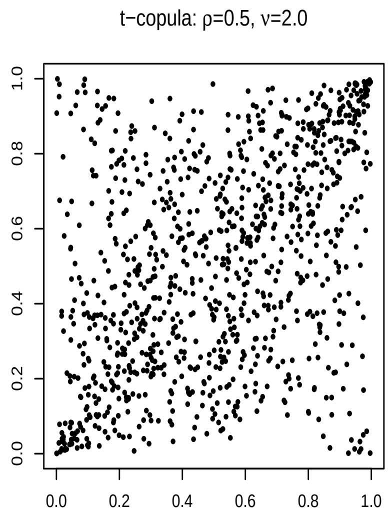
$$C_{\nu,\rho}(u_1, u_2) = \int_0^{u_1}\int_0^{u_2} \frac{\frac{\nu}{2}\,\Gamma\left(\frac{\nu}{2}\right)^2 \left(1 + \frac{t_\nu^{-1}(v_1)^2 + t_\nu^{-1}(v_2)^2 - 2\,\rho\, t_\nu^{-1}(v_1)\, t_\nu^{-1}(v_2)}{\nu\,(1-\rho^2)}\right)^{-\frac{\nu+2}{2}}}{\sqrt{1-\rho^2}\,\Gamma\left(\frac{\nu+1}{2}\right)^2 \left(\left(1 + \frac{t_\nu^{-1}(v_1)^2}{\nu}\right)\left(1 + \frac{t_\nu^{-1}(v_2)^2}{\nu}\right)\right)^{-\frac{\nu+1}{2}}}\, dv_2\, dv_1,$$

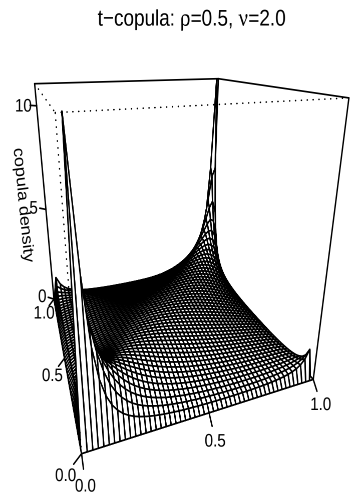
where $t_\nu(x) := \int_{-\infty}^{x} (1 + y^2/\nu)^{-(\nu+1)/2}\, dy\, \Gamma((\nu+1)/2)/\sqrt{\nu\,\pi}/\Gamma(\nu/2)$ is the distribution function of a univariate t-distribution with ν degrees of freedom.

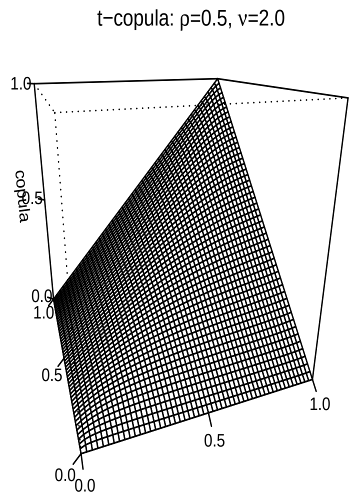
***Observation**:*

- *This family of copulas is two-parametric.*
- *For every ν and ρ, $C_{\nu,\rho}$ is an absolutely continuous copula.*

t-copulas







A scatter plot and the density of a bivariate t-copula, as well as the bivariate t-copula itself.

t-copulas

Remark 15 (Important stylized facts of the bivariate t-copula)

(a) ***Symmetries**: Like the bivariate Gaussian copula, the bivariate t-copula is both radially symmetric and exchangeable.*

(b) ***Concordance measures**: Kendall's τ and Spearman's ρ_S are the same as for the bivariate Gaussian copula, independent of the degrees of freedom ν.*

$$\tau = \frac{2}{\pi}\arcsin(\rho), \qquad \text{(Kendall's } \tau\text{)}$$

$$\rho_S = \frac{6}{\pi}\arcsin(\rho/2). \qquad \text{(Spearman's } \rho_S\text{)}$$

(c) ***Tail dependence**: Unlike in the case of the Gaussian copula, the lower- and upper-tail dependence coefficients are not zero. They are given by*

$$UTD_{\nu,\rho} = LTD_{\nu,\rho} = 2 \cdot t_{\nu+1}\left(-\sqrt{\frac{(\nu+1)(1-\rho)}{1+\rho}}\right).$$

Lecture III

Estimation of copulas: parametric and semiparametric approaches

Setup

- Let $d = 2$, i.e. bivariate case
- Let $C(u_1, u_2; \boldsymbol{\theta})$ be a family of copulas the parameter vector $\boldsymbol{\theta}$
- Let $C(u_1, u_2; \boldsymbol{\theta})$ be absolutely continuous
- Let a sample from the random vector (X_1, X_2) be given

$\rightarrow$ Focus on estimation of $\boldsymbol{\theta}$

Specification

- α_k $(k = 1, 2)$ denotes the parameter vector of the marginal distribution and θ denotes the parameter vector of the copula
- Let (X_1, X_2) denote a continuous bivariate random variable and $F_k(x; \alpha_k)$ and $f_k(x; \alpha_k)$ be the *cdf* and the *pdf* of X_k
- Let $U_k = F_k(X_k; \alpha_k)$
- $C(u_1, u_2; \theta)$ denotes the joint *cdf* of (U_1, U_2), $c(u_1, u_2; \theta)$ denotes the *pdf* corresponding to $C(u_1, u_2; \theta)$
- $H(x_1, x_2; \eta)$ and $h(x_1, x_2; \eta)$ denote the *cdf* and *pdf* of (X_1, X_2), respectively, where $\eta = (\alpha_1', \alpha_2', \theta')'$

$\rightarrow$ Estimation of θ using iid observations (x_{1i}, x_{2i}), $i = 1, \ldots, n$

Estimation methods: Overview

- Maximum Likelihood Estimation (MLE)
- Inference Function for Margins method (IFM)
- Semiparametric method (SP) / Pseudo Maximum Likelihood Estimation (PMLE)

Maximum likelihood estimation

The joint density function $h(x_1, x_2; \boldsymbol{\eta})$ of (X_1, X_2) can be expressed as follows:

$$h(x_1, x_2; \boldsymbol{\eta}) = c(F_1(x_1; \boldsymbol{\alpha}_1), F_2(x_2; \boldsymbol{\alpha}_2); \boldsymbol{\theta}) f_1(x_1; \boldsymbol{\alpha}_1) f_2(x_2; \boldsymbol{\alpha}_2)$$

Therefore, the log-likelihood function takes the form:

$$L(\boldsymbol{\eta}) = \sum_{i=1}^{n} log[c(F_1(x_{1i}; \boldsymbol{\alpha}_1), F_2(x_{2i}; \boldsymbol{\alpha}_2); \boldsymbol{\theta}) f_1(x_{1i}; \boldsymbol{\alpha}_1) f_2(x_{2i}; \boldsymbol{\alpha}_2)]$$

- MLE of $\boldsymbol{\eta}$: $\widehat{\boldsymbol{\eta}}^{MLE} := (\widehat{\boldsymbol{\alpha}}_1', \widehat{\boldsymbol{\alpha}}_2', \widehat{\boldsymbol{\theta}}')' = \underset{\eta}{argmax}\ L(\boldsymbol{\eta})$
- Under some regularity assumptions, we get $\widehat{\boldsymbol{\eta}}^{MLE}$ from solving:

$$(\partial L / \partial \boldsymbol{\alpha}_1^T, \partial L / \partial \boldsymbol{\alpha}_2^T, \partial L / \partial \boldsymbol{\theta}^T)^T = \mathbf{0}$$

- $\sqrt{n}(\widehat{\boldsymbol{\eta}}^{MLE} - \boldsymbol{\eta}) \xrightarrow{d} N(\mathbf{0}, \mathcal{I}(\boldsymbol{\eta})^{-1})$, for $n \to \infty$, where $\mathcal{I}(\boldsymbol{\eta}) = \mathcal{I}$ is the Fisher information matrix
- MLE is asymptotically efficient and hence is the preferred first option, when the model is correctly specified

Inference for margins

Problems with MLE:
One does not usually have closed form estimators and numerical techniques are needed. For MLE, the number of parameters increases with the dimension and numerical optimization becomes more time consuming.

Solution: Two-stage estimation

1. Each marginal distribution is estimated separately: Marginal log-likelihoods $L_k(\boldsymbol{\alpha}_k) = \sum_{i=1}^{n} log(f_k(x_{ik}; \boldsymbol{\alpha}_k))$, $k = 1, 2$ are separately maximized to get $\widehat{\boldsymbol{\alpha}}_{\mathbf{1}}^{IFM}$, $\widehat{\boldsymbol{\alpha}}_{\mathbf{2}}^{IFM}$
2. $\boldsymbol{\theta}$ is estimated by substituting $\widehat{\boldsymbol{\alpha}}_{\boldsymbol{k}}^{IFM}$ for $\boldsymbol{\alpha}_k$ in the log-likelihood function for the joint distribution and then maximizing the resulting function

Inference for margins

Thus, the IFM estimate $\widehat{\boldsymbol{\theta}}^{IMF}$ of $\boldsymbol{\theta}$ is the maximum of

$$\tilde{L}(\boldsymbol{\theta}) = \sum_{i=1}^{n} log[c(F_1(x_{1i}; \widehat{\boldsymbol{\alpha}}_{\mathbf{1}}{}^{IFM}), F_2(x_{2i}; \widehat{\boldsymbol{\alpha}}_{\mathbf{2}}{}^{IFM}); \boldsymbol{\theta})]$$

Under some regularity assumptions, $\widehat{\boldsymbol{\eta}}^{IFM}$ is the solution of

$$(\partial L_1/\partial \boldsymbol{\alpha}_1^T, \partial L_2/\partial \boldsymbol{\alpha}_2^T, \partial \tilde{L}/\partial \boldsymbol{\theta}^T)^T = \mathbf{0}$$

This procedure is computationally simpler than estimating all parameters $\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \boldsymbol{\theta}$ simultaneously.

Inference for margins: Asymptotic normality

From the theory of inference functions,

$$\sqrt{n}(\widehat{\boldsymbol{\eta}}^{IFM} - \boldsymbol{\eta}) \xrightarrow{d} N(\mathbf{0}, V), \quad n \to \infty$$

The asymptotic covariance matrix for $\widehat{\boldsymbol{\eta}}^{IFM}$ is

$$V = (-D_g^{-1}) M_g (-D_g^{-1})^T,$$

where $M_g = Cov(\mathbf{g}(\mathbf{X}; \boldsymbol{\eta})) = E[\mathbf{g}\mathbf{g}^T]$,
$D_g = E[\partial \mathbf{g}(\mathbf{X}, \boldsymbol{\eta}) / \partial \boldsymbol{\eta}^T]$,
$\mathbf{X} = (X_1, X_2)$,
$\mathbf{g}^T = (g_1^T, g_2^T, g_3^T)$,
$g_k = \partial l_k / \partial \boldsymbol{\alpha}_k$, $l_k = log\ f_k(\cdot; \boldsymbol{\alpha}_k)$ for $k = 1, 2$,
$g_3 = \partial l / \partial \boldsymbol{\theta}$, $l = log\ h(\cdot, \cdot; \boldsymbol{\eta})$.

[Joe, (2005)]

Inference for margins: Asymptotic relative efficiency

- Comparison of the ML estimator and the IFM estimator for **scalar** θ in terms of their variances
- The ratio of the variance of the first estimator to the variance of the second estimator is called the asymptotic efficiency of the second estimator with respect to the first
- Numerical computations showed that the **IFM has good efficiency**
- IFM estimator $\widehat{\theta}^{IMF}$ has very high efficiency
- However, in cases of extreme dependence near the Frèchet bounds there can be a loss of efficiency of the univariate parameter estimators $\widehat{\alpha}_1{}^{IFM}$, $\widehat{\alpha}_2{}^{IFM}$

The ML and IFM methods are **completely parametric** because they require the model to be specified up to a finite number of unknown parameters. A possible shortcoming of these two methods of estimating θ is that they are **likely to be inconsistent** even if just one marginal distribution is misspecified.

Solution: Semiparametric method

Semiparametric method

- Marginal distributions are allowed to have arbitrary and unknown functional forms
- Two-stage estimation as in IFM
- Difference: The marginal distributions are estimated nonparametrically by **their sample empirical distributions**
- More specifically: Let $\hat{F}_k$ denote the *rescaled empirical cdf* of $x_{k1}, \ldots, x_{kn}$, $(k = 1, 2)$, defined as

$$\hat{F}_k(x) = \frac{1}{n+1} \sum_{i=1}^{n} I(x_{ki} \leq x)$$

- Rescaling the *ecdf* with $\frac{n}{n+1}$ ensures that the first order condition of the log-likelihood function for the joint distribution is well defined for all finite n

Semiparametric method: Pseudo log-likelihood

Thus, the SP estimate (PMLE) $\tilde{\boldsymbol{\theta}}^{SP}$ of $\boldsymbol{\theta}$ is the maximum of the **pseudo log-likelihood**

$$\hat{L}(\boldsymbol{\theta}) = \sum_{i=1}^{n} log[c(\hat{F}_1(x_{1i}), \hat{F}_2(x_{2i}); \boldsymbol{\theta}]$$

Proposition 3

Let $R_n := \frac{1}{n}\sum_{i=1}^{n} J(\hat{F}_1(x_{1i}), \hat{F}_2(x_{2i}))$ *and* $J(u_1, u_2)$ *be a continuous function from* $(0,1)^2$ *into* $\mathbb{R}$ *such that*

$$\mu := E[J(F_1(X_1), F_2(X_2))] = \int J(u_1, u_2) dC(u_1, u_2)$$

exists. Further, let J admit continuous partial derivatives $J_i(u_1, u_2) = \partial J / \partial u_i$ *for* $i = 1, 2$*. Under suitable regularity conditions, it follows that*
(i) $R_n \to \mu$ *almost surely.*
(ii) $\sqrt{n}(R_n - \mu) \to N(0, \sigma^2)$ *in distribution, where*

$$\sigma^2 := Var[J(F_1(X_1), F_2(X_2)) + \sum_{i=1}^{2} \int \mathbf{1}_{\{}X_i \leq x_i\} J_i(F_1(x_1), F_2(x_2)) dH(x_1, x_2)].$$

Proof: [Genest, (1995)]

Semiparametric method: Asymptotic normality

Let $l(u_1, u_2, \theta) = log\ c(u_1, u_2, \theta)$ and use indices $1, 2$ and θ to denote partial derivatives of l with respect to u_1, u_2 and θ respectively.

Proposition 4

Under suitable regularity conditions, the semiparametric estimator $\tilde{\theta}^{SP}$ ***is consistent*** *and $\sqrt{n}(\tilde{\theta}^{SP} - \theta)$* ***is asymptotically normal*** *with variance $\nu^2 = \sigma^2/\beta^2$, where*

$$\sigma^2 := Var[l_\theta(F_1(X_1), F_2(X_2), \theta) + W_1(X_1) + W_2(X_2)],$$

$$W_i(X_i) := \int \mathbf{1}\{F_i(X_i) \leq u_i\} l_{\theta,i}(u_1, u_2, \theta) c(u_1, u_2, \theta) du_1 du_2,$$

$$\beta := -E[l_{\theta,\theta}(F_1(X_1), F_2(X_2), \theta)] = E[l_\theta^2(F_1(X_1), F_2(X_2), \theta)].$$

Semiparametric method: Asymptotic variance

Note that

$$\sigma^2 = \beta + Var[W_1(X_1) + W_2(X_2)].$$

Therefore, it follows that

$$\nu^2 = \frac{\sigma^2}{\beta^2} = \frac{1}{\beta} + \frac{Var[W_1(X_1) + W_2(X_2)]}{\beta^2} \geq \frac{1}{\beta}$$

- The inequality expresses the fact that $\tilde{\boldsymbol{\theta}}^{SP}$ has a larger asymptotic variance than the MLE $\widehat{\boldsymbol{\theta}}^{MLE}$ of $\boldsymbol{\theta}$ computed under the assumption that the marginals are known
- Equality in the above inequality occurs when the copula approaches the independence copula (with paramter $\boldsymbol{\theta}_{\Pi_2}$)

Lecture IV

Vine copulas

Multivariate copulas

Elliptical copulas

- Many parameters (correlation matrix).
- Only symmetric dependence.
- Student's t copula: only one degrees of freedom parameter.

Archimedean copulas

- Few parameters (usually one or two).
- Same dependence for all pairs.

Solution: Pair-copula constructions or vine copulas

Pair-copula construction in 3 dimensions

Factorization

$$f(x_1, x_2, x_3) = f_{3|12}(x_3|x_1, x_2) f_{2|1}(x_2|x_1) f_1(x_1)$$

Pair-copula construction in 3 dimensions

Factorization

$$f(x_1, x_2, x_3) = f_{3|12}(x_3|x_1, x_2) f_{2|1}(x_2|x_1) f_1(x_1)$$

Using Sklar's Theorem for $f_{12}(x_1, x_2)$, $f_{13|2}(x_1, x_3|x_2)$ and $f_{23}(x_2, x_3)$ implies

$$\begin{aligned}
f_{3|12}(x_3|x_1, x_2) &= f_{13|2}(x_1, x_3|x_2) \frac{1}{f_{1|2}(x_1|x_2)} \\
&= c_{13;2}(F_{1|2}(x_1|x_2), F_{3|2}(x_3|x_2); x_2) \cancel{f_{1|2}(x_1|x_2)} \frac{f_{3|2}(x_3|x_2)}{\cancel{f_{1|2}(x_1|x_2)}} \\
&= c_{13;2}(F_{1|2}(x_1|x_2), F_{3|2}(x_3|x_2); x_2) \underbrace{f_{3|2}(x_3|x_2)} \\
&= c_{13;2}(F_{1|2}(x_1|x_2), F_{3|2}(x_3|x_2); x_2) c_{23}(F_2(x_2), F_3(x_3)) f_3(x_3)
\end{aligned}$$

3-dimensional pair-copula construction

$$\begin{aligned}
f(x_1, x_2, x_3) = {} & c_{13;2}(F_{1|2}(x_1|x_2), F_{3|2}(x_3|x_2); x_2) c_{23}(F_2(x_2), F_3(x_3)) \\
& \times \; c_{12}(F_1(x_1), F_2(x_2)) f_3(x_3) f_2(x_2) f_1(x_1)
\end{aligned}$$

Pair-copula construction in 3 dimensions

Factorization

$$f(x_1, x_2, x_3) = f_{3|12}(x_3|x_1, x_2) f_{2|1}(x_2|x_1) f_1(x_1)$$

Using Sklar's Theorem for $f_{12}(x_1, x_2)$, $f_{13|2}(x_1, x_3|x_2)$ and $f_{23}(x_2, x_3)$ implies

$$\begin{aligned}
&= c_{13;2}(F_{1|2}(x_1|x_2), F_{3|2}(x_3|x_2); x_2) \cancel{f_{1|2}(x_1|x_2)} \frac{f_{3|2}(x_3|x_2)}{\cancel{f_{1|2}(x_1|x_2)}} \\
&= c_{13;2}(F_{1|2}(x_1|x_2), F_{3|2}(x_3|x_2); x_2) \underbrace{f_{3|2}(x_3|x_2)} \\
&= c_{13;2}(F_{1|2}(x_1|x_2), F_{3|2}(x_3|x_2); x_2) c_{23}(F_2(x_2), F_3(x_3)) f_3(x_3)
\end{aligned}$$

3-dimensional pair-copula construction

$$\begin{aligned}
f(x_1, x_2, x_3) = {} & c_{13;2}(F_{1|2}(x_1|x_2), F_{3|2}(x_3|x_2); x_2) c_{23}(F_2(x_2), F_3(x_3)) \\
& \times \; c_{12}(F_1(x_1), F_2(x_2)) f_3(x_3) f_2(x_2) f_1(x_1)
\end{aligned}$$

Pair-copula construction in 3 dimensions

Factorization

$$f(x_1, x_2, x_3) = f_{3|12}(x_3|x_1, x_2)f_{2|1}(x_2|x_1)f_1(x_1)$$

Using Sklar's Theorem for $f_{12}(x_1, x_2)$, $f_{13|2}(x_1, x_3|x_2)$ and $f_{23}(x_2, x_3)$ implies

$$= c_{13;2}(F_{1|2}(x_1|x_2), F_{3|2}(x_3|x_2); x_2)c_{23}(F_2(x_2), F_3(x_3))f_3(x_3)$$

3-dimensional pair-copula construction

$$\begin{aligned} f(x_1, x_2, x_3) = {} & c_{13;2}(F_{1|2}(x_1|x_2), F_{3|2}(x_3|x_2); x_2)c_{23}(F_2(x_2), F_3(x_3)) \\ & \times \; c_{12}(F_1(x_1), F_2(x_2))f_3(x_3)f_2(x_2)f_1(x_1) \end{aligned}$$

Pair-copula construction in 3 dimensions

Factorization

$$f(x_1, x_2, x_3) = f_{3|12}(x_3|x_1, x_2) f_{2|1}(x_2|x_1) f_1(x_1)$$

Using Sklar's Theorem for $f_{12}(x_1, x_2)$, $f_{13|2}(x_1, x_3|x_2)$ and $f_{23}(x_2, x_3)$ implies

$$\begin{aligned}
f_{2|1}(x_2|x_1) &= c_{12}(F_1(x_1), F_2(x_2)) f_2(x_2) \\
f_{3|12}(x_3|x_1, x_2) &= f_{13|2}(x_1, x_3|x_2) \frac{1}{f_{1|2}(x_1|x_2)} \\
&= c_{13;2}(F_{1|2}(x_1|x_2), F_{3|2}(x_3|x_2); x_2) \cancel{f_{1|2}(x_1|x_2)} \frac{f_{3|2}(x_3|x_2)}{\cancel{f_{1|2}(x_1|x_2)}} \\
&= c_{13;2}(F_{1|2}(x_1|x_2), F_{3|2}(x_3|x_2); x_2) \underbrace{f_{3|2}(x_3|x_2)} \\
&= c_{13;2}(F_{1|2}(x_1|x_2), F_{3|2}(x_3|x_2); x_2) c_{23}(F_2(x_2), F_3(x_3)) f_3(x_3)
\end{aligned}$$

Pair-copula construction in 3 dimensions

Factorization

$$f(x_1, x_2, x_3) = f_{3|12}(x_3|x_1, x_2) f_{2|1}(x_2|x_1) f_1(x_1)$$

Using Sklar's Theorem for $f_{12}(x_1, x_2)$, $f_{13|2}(x_1, x_3|x_2)$ and $f_{23}(x_2, x_3)$ implies

$$\begin{aligned}
f_{2|1}(x_2|x_1) &= c_{12}(F_1(x_1), F_2(x_2)) f_2(x_2) \\
f_{3|12}(x_3|x_1, x_2) &= f_{13|2}(x_1, x_3|x_2) \frac{1}{f_{1|2}(x_1|x_2)} \\
&= c_{13;2}(F_{1|2}(x_1|x_2), F_{3|2}(x_3|x_2); x_2) \cancel{f_{1|2}(x_1|x_2)} \frac{f_{3|2}(x_3|x_2)}{\cancel{f_{1|2}(x_1|x_2)}} \\
&= c_{13;2}(F_{1|2}(x_1|x_2), F_{3|2}(x_3|x_2); x_2) \underbrace{f_{3|2}(x_3|x_2)} \\
&= c_{13;2}(F_{1|2}(x_1|x_2), F_{3|2}(x_3|x_2); x_2) c_{23}(F_2(x_2), F_3(x_3)) f_3(x_3)
\end{aligned}$$

3-dimensional pair-copula construction

$$\begin{aligned}
f(x_1, x_2, x_3) = {} & c_{13;2}(F_{1|2}(x_1|x_2), F_{3|2}(x_3|x_2); x_2) c_{23}(F_2(x_2), F_3(x_3)) \\
& \times \; c_{12}(F_1(x_1), F_2(x_2)) f_3(x_3) f_2(x_2) f_1(x_1)
\end{aligned}$$

Pair-copula construction in d dimensions

Factorization

$$f(x_1, \ldots, x_d) = \left[\prod_{k=2}^{d} f_{k|1,\ldots,k-1}(x_k|x_1, \ldots, x_{k-1})\right] \times f_1(x_1)$$

Pair-copula construction in d dimensions

Factorization

$$f(x_1, \ldots, x_d) = \left[\prod_{k=2}^{d} f_{k|1,\ldots,k-1}(x_k | x_1, \ldots, x_{k-1})\right] \times f_1(x_1)$$

For distinct $i, j, i_1, \ldots, i_k$ with $i < j$ and $i_1 < \ldots < i_k$ let

$$c_{i,j;i_1,\ldots,i_k} := c_{i,j;i_1,\ldots,i_k}(F_{i|i_1,\ldots,i_k}(x_i | x_{i_1}, \ldots, x_{i_k}), (F_{j|i_1,\ldots,i_k}(x_j | x_{i_1}, \ldots, x_{i_k})).$$

Then $f_{k|1,\ldots,k-1}(x_k | x_1, \ldots, x_{k-1}) = \left[\prod_{\ell=1}^{k-2} c_{\ell,k;\ell+1,\ldots,k-1}\right] \times c_{k-1,k} \times f_k(x_k)$

Pair-copula construction in d dimensions

Factorization

$$f(x_1, \ldots, x_d) = \left[\prod_{k=2}^{d} f_{k|1,\ldots,k-1}(x_k|x_1, \ldots, x_{k-1})\right] \times f_1(x_1)$$

For distinct $i, j, i_1, \ldots, i_k$ with $i < j$ and $i_1 < \ldots < i_k$ let

$$c_{i,j;i_1,\ldots,i_k} := c_{i,j;i_1,\ldots,i_k}\left(F_{i|i_1,\ldots,i_k}(x_i|x_{i_1}, \ldots, x_{i_k}), \left(F_{j|i_1,\ldots,i_k}(x_j|x_{i_1}, \ldots, x_{i_k}\right)\right).$$

Then $f_{k|1,\ldots,k-1}(x_k|x_1,\ldots,x_{k-1}) = \left[\prod_{\ell=1}^{k-2} c_{\ell,k;\ell+1,\ldots,k-1}\right] \times c_{k-1,k} \times f_k(x_k)$

With $\ell = i$ and $k = i + j$ it follows that:

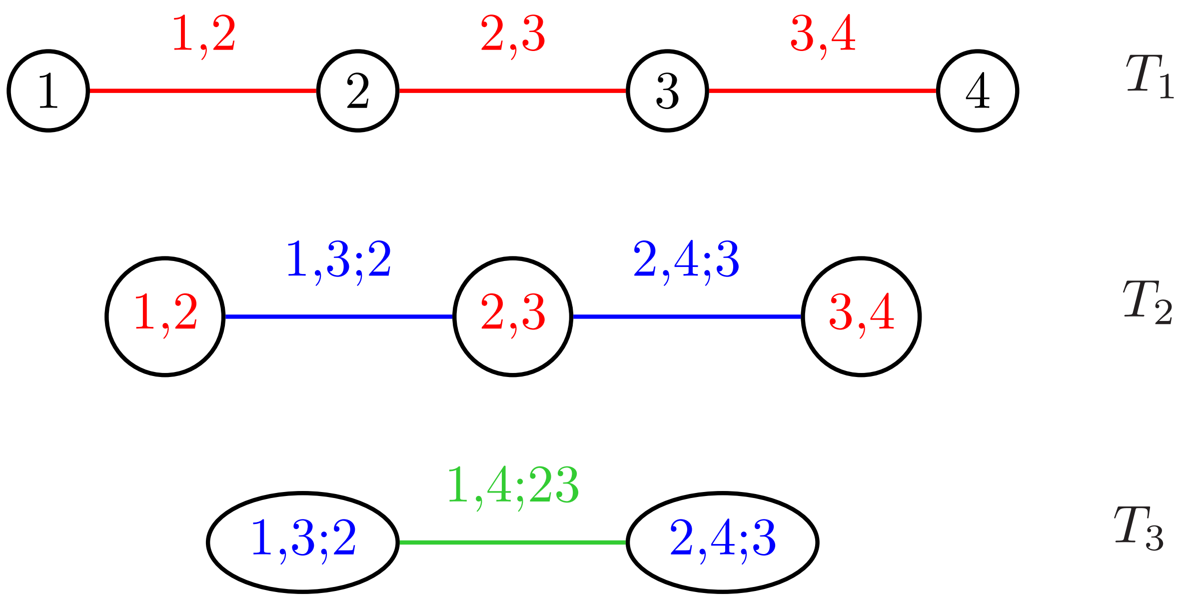
d-dimensional pair-copula construction

$$f(x_1, \ldots, x_d) = \left[\prod_{j=1}^{d-1}\prod_{i=1}^{d-j} c_{i,i+j;i+1,\ldots,i+j-1}\right] \times \left[\prod_{k=1}^{d} f_k(x_k)\right]$$

4-dimensional pair-copula construction

4-dimensional pair-copula construction

$$f = f_1 \cdot f_2 \cdot f_3 \cdot f_4 \cdot c_{12} \cdot c_{23} \cdot c_{34} \cdot c_{13;2} \cdot c_{24;3} \cdot c_{14;23}$$



Some graph theory

Decomposition into pair-copulas is
not unique.

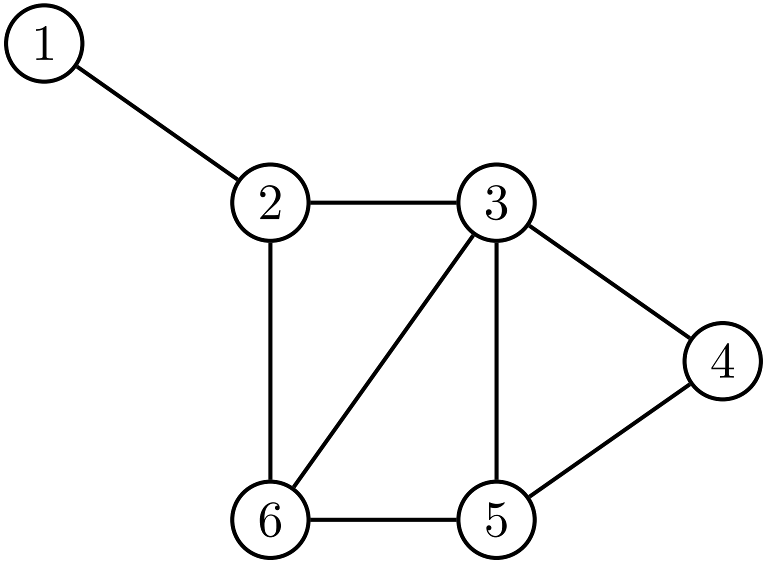
- Graph-theoretical model
  to organize pair-copula constructions.

Some graph theory

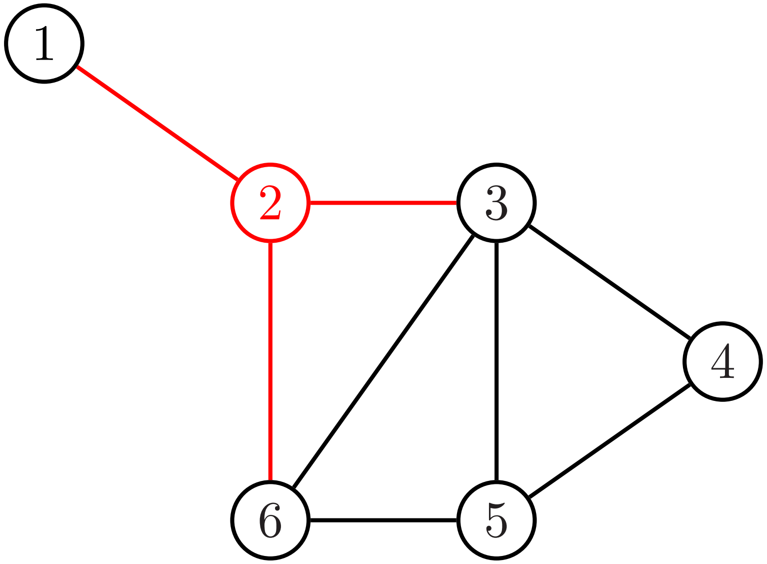
Decomposition into pair-copulas is
not unique.

- Graph-theoretical model
  to organize pair-copula constructions.

- Graph $G = (N, E)$ with
  - nodes N and
  - (undirected) edges $E \subset N \times N$.



Some graph theory



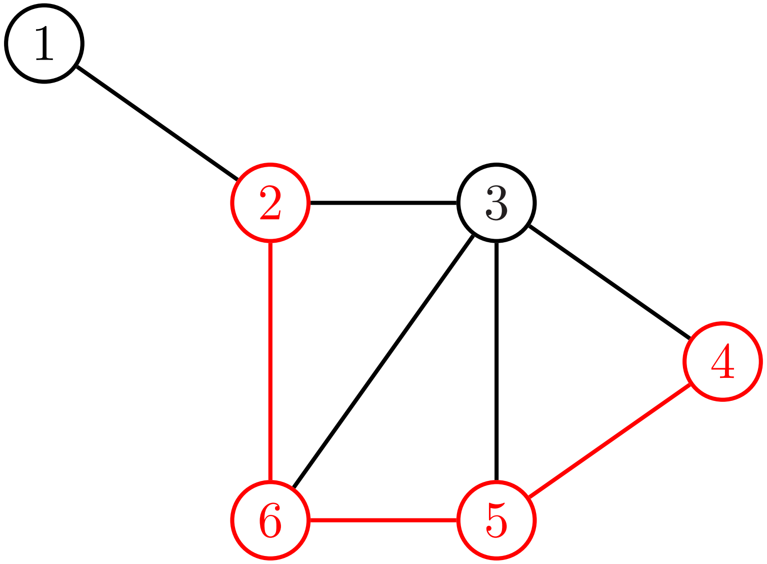
Decomposition into pair-copulas is not unique.

- Graph-theoretical model
  to organize pair-copula constructions.

- Graph $G = (N, E)$ with
  - nodes N and
  - (undirected) edges $E \subset N \times N$.

- Degree of a node: number of nodes connected to this node.

Some graph theory



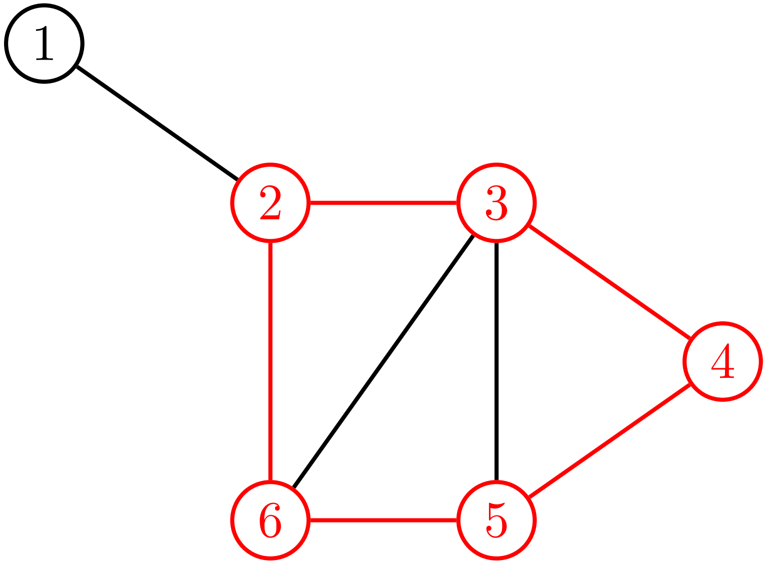
Decomposition into pair-copulas is not unique.

- ▶ Graph-theoretical model to organize pair-copula constructions.

- Graph $G = (N, E)$ with
  - nodes N and
  - (undirected) edges $E \subset N \times N$.

- Degree of a node: number of nodes connected to this node.
- Path: sequence of connected nodes.

Some graph theory



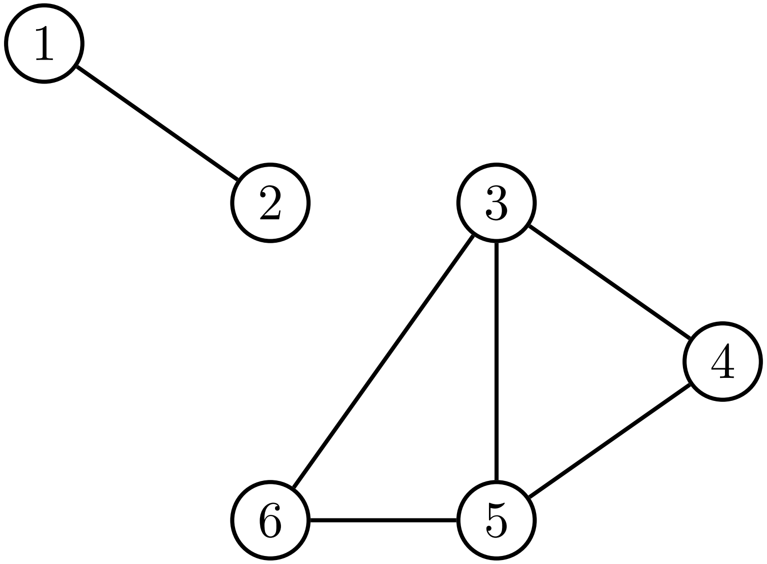
Decomposition into pair-copulas is not unique.

- ▶ Graph-theoretical model to organize pair-copula constructions.

- Graph $G = (N, E)$ with
  - nodes N and
  - (undirected) edges $E \subset N \times N$.

- Degree of a node: number of nodes connected to this node.
- Path: sequence of connected nodes.
- Cycle: path with end node = start node.

Some graph theory



Decomposition into pair-copulas is not unique.

- Graph-theoretical model to organize pair-copula constructions.

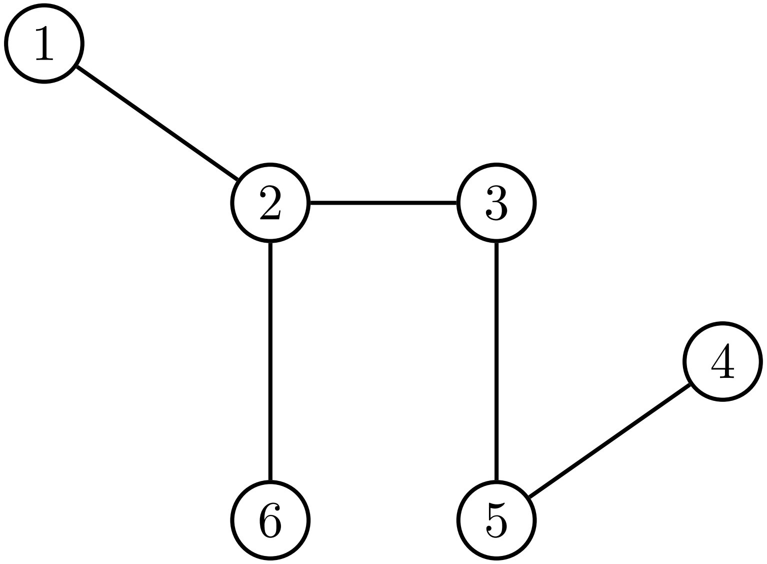
- Graph $G = (N, E)$ with
  - nodes N and
  - (undirected) edges $E \subset N \times N$.

- Degree of a node: number of nodes connected to this node.
- Path: sequence of connected nodes.
- Cycle: path with end node = start node.
- Connected graph: path from each node to each other node.

Some graph theory

Decomposition into pair-copulas is not unique.

- Graph-theoretical model to organize pair-copula constructions.



- Graph $G = (N, E)$ with
  - nodes N and
  - (undirected) edges $E \subset N \times N$.

- Degree of a node: number of nodes connected to this node.
- Path: sequence of connected nodes.
- Cycle: path with end node = start node.
- Connected graph: path from each node to each other node.
- Tree: connected, acyclic graph.

Regular vines

Bedford and Cooke (2002) introduced the graphical model called regular vine to organize pair-copula constructions.

Regular vines

Bedford and Cooke (2002) introduced the graphical model called regular vine to organize pair-copula constructions.

Regular vine (R-vine)

A d-dimensional regular vine is a linked sequence of $d-1$ trees $T_1, \ldots, T_{d-1}$ with edge sets $E_1, \ldots, E_{d-1}$.

Regular vines

Bedford and Cooke (2002) introduced the graphical model called regular vine to organize pair-copula constructions.

Regular vine (R-vine)

A d-dimensional regular vine is a linked sequence of $d-1$ trees $T_1, \ldots, T_{d-1}$ with edge sets $E_1, \ldots, E_{d-1}$.

Example ($d=3$):

Regular vines

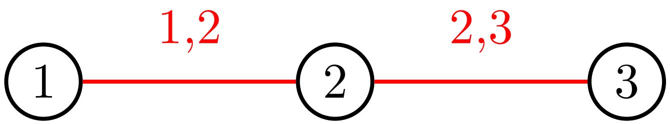
Bedford and Cooke (2002) introduced the graphical model called regular vine to organize pair-copula constructions.

Regular vine (R-vine)

A d-dimensional regular vine is a linked sequence of $d-1$ trees $T_1, \ldots, T_{d-1}$ with edge sets $E_1, \ldots, E_{d-1}$.

1. Tree j has $d+1-j$ nodes and $d-j$ edges.

Example ($d = 3$):



Regular vines

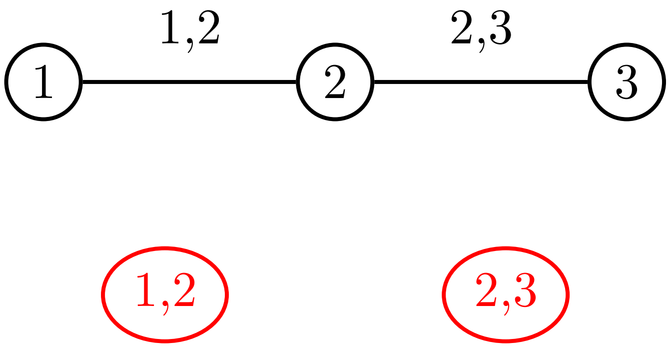
Bedford and Cooke (2002) introduced the graphical model called regular vine to organize pair-copula constructions.

Regular vine (R-vine)

A d-dimensional regular vine is a linked sequence of $d-1$ trees $T_1, \ldots, T_{d-1}$ with edge sets $E_1, \ldots, E_{d-1}$.

1. Tree j has $d+1-j$ nodes and $d-j$ edges.
2. Edges in tree j become nodes in tree $j+1$.

Example ($d=3$):



Regular vines

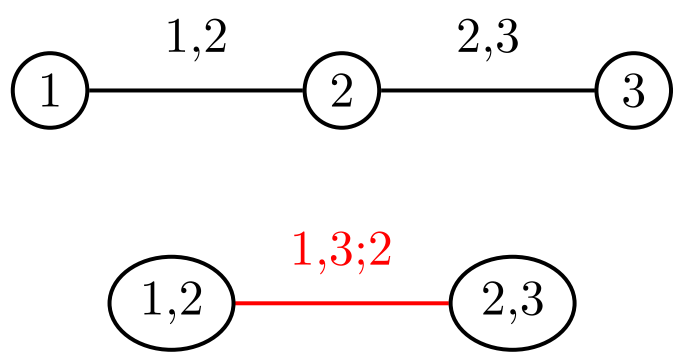
Bedford and Cooke (2002) introduced the graphical model called regular vine to organize pair-copula constructions.

Regular vine (R-vine)

A d-dimensional regular vine is a linked sequence of $d-1$ trees $T_1, \dots, T_{d-1}$ with edge sets $E_1, \dots, E_{d-1}$.

1. Tree j has $d+1-j$ nodes and $d-j$ edges.
2. Edges in tree j become nodes in tree $j+1$.
3. Proximity condition: Two nodes in tree $j+1$ are joined by an edge only if the corresponding edges in tree j share a node.

Example ($d=3$):



Regular vine distributions and copulas

Regular vine distribution

A *d*-dimensional regular vine distribution has the following components:

- A regular vine tree structure.
- Each edge corresponds to a pair-copula density.
- The density of a regular vine distribution is defined by
  - the product of pair-copula densities over the $d(d-1)/2$ edges identified by the regular vine trees and
  - the product of the marginal densities.

Regular vine distributions and copulas

Regular vine distribution

A *d*-dimensional regular vine distribution has the following components:

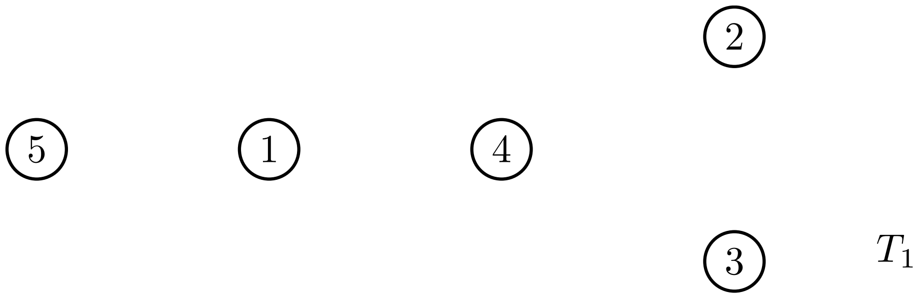
- A regular vine tree structure.
- Each edge corresponds to a pair-copula density.
- The density of a regular vine distribution is defined by
  - the product of pair-copula densities over the $d(d-1)/2$ edges identified by the regular vine trees and
  - the product of the marginal densities.

A regular vine copula is defined as the product of pair-copulas determined through a regular vine.

Example ($d = 3$):

$$
\begin{aligned}
f(x_1, x_2, x_3) = {} & c_{13;2}(F_{1|2}(x_1|x_2), F_{3|2}(x_3|x_2); x_2) c_{23}(F_2(x_2), F_3(x_3)) \\
& \times \; c_{12}(F_1(x_1), F_2(x_2)) f_3(x_3) f_2(x_2) f_1(x_1)
\end{aligned}
$$

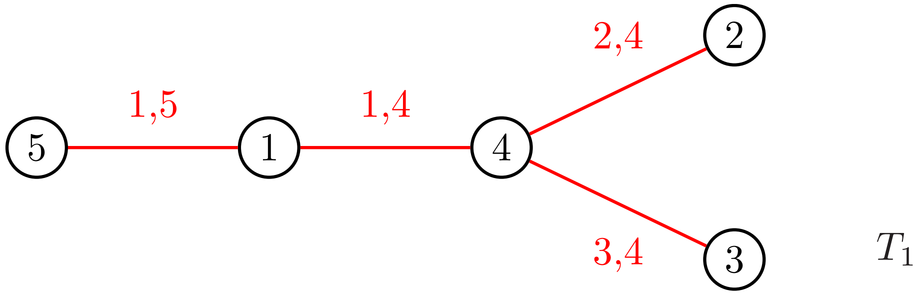
Example of a five-dimensional R-vine



Density

$$
\begin{aligned}
f = f_1 \cdot f_2 \cdot f_3 \cdot f_4 \cdot f_5 \\
\cdot c_{14} \cdot c_{15} \cdot c_{24} \cdot c_{34} \\
\cdot c_{12;4} \cdot c_{13;4} \cdot c_{45;1} \\
\cdot c_{23;14} \cdot c_{35;14} \\
\cdot c_{25;134}
\end{aligned}
$$

Example of a five-dimensional R-vine



Density

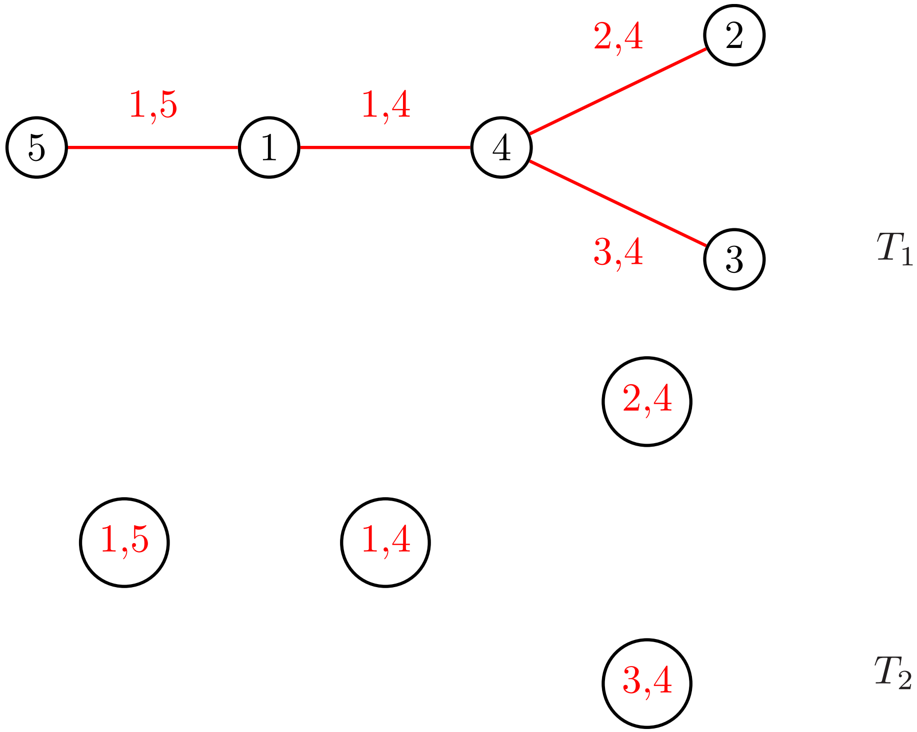
$$f = f_1 \cdot f_2 \cdot f_3 \cdot f_4 \cdot f_5$$

$$\cdot\, c_{12;4} \cdot c_{13;4} \cdot c_{45;1}$$

$$\cdot\, c_{23;14} \cdot c_{35;14}$$

$$\cdot\, c_{25;134}$$

Example of a five-dimensional R-vine



Density

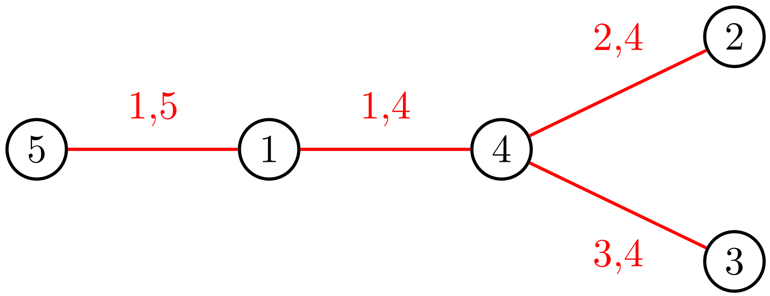
$$f = f_1 \cdot f_2 \cdot f_3 \cdot f_4 \cdot f_5$$

$$\cdot\, c_{12;4} \cdot c_{13;4} \cdot c_{45;1}$$

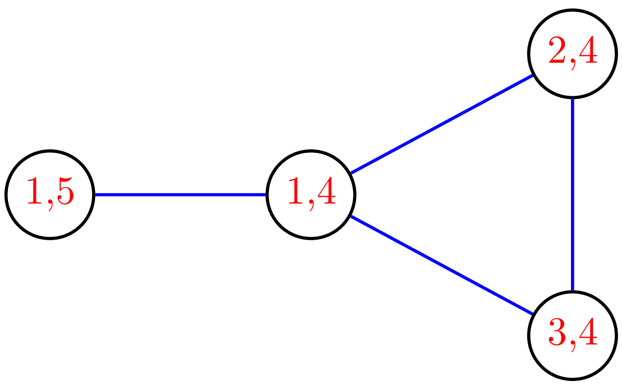
$$\cdot\, c_{23;14} \cdot c_{35;14}$$

$$\cdot\, c_{25;134}$$

Example of a five-dimensional R-vine



T_1



T_2

Density

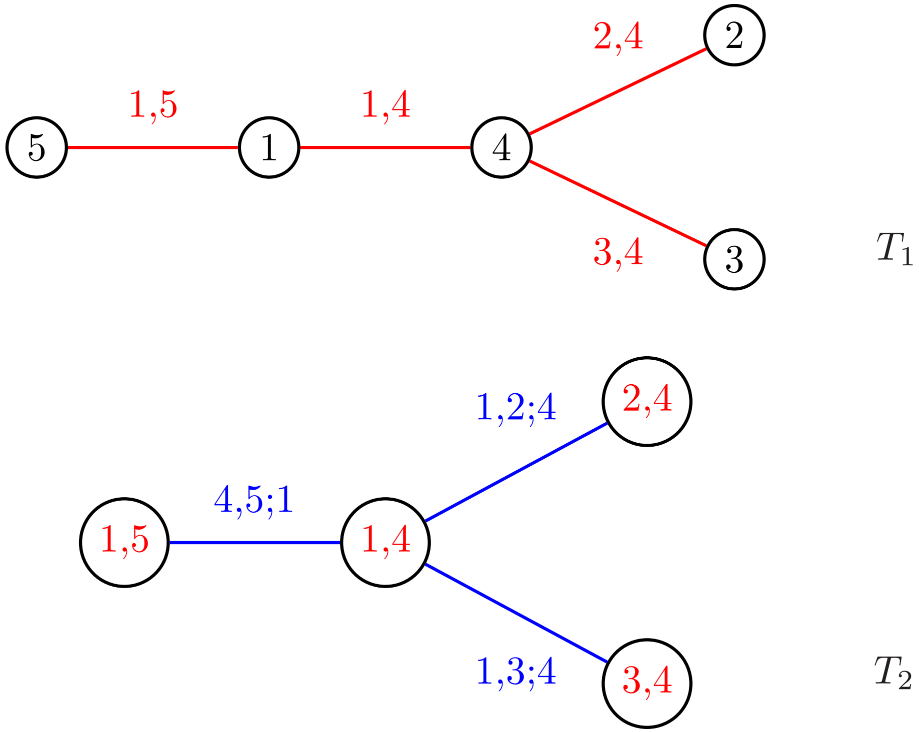
$$f = f_1 \cdot f_2 \cdot f_3 \cdot f_4 \cdot f_5$$

$$\cdot\, c_{12;4} \cdot c_{13;4} \cdot c_{45;1}$$

$$\cdot\, c_{23;14} \cdot c_{35;14}$$

$$\cdot\, c_{25;134}$$

Example of a five-dimensional R-vine



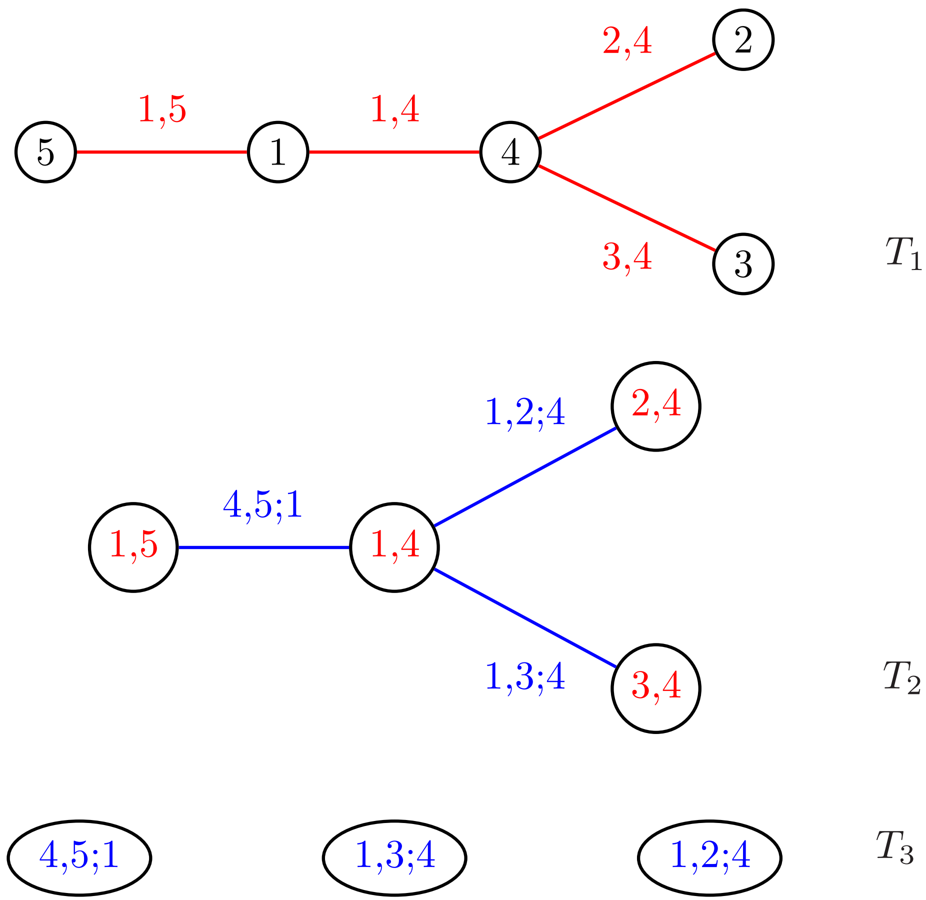
Density

$$f = f_1 \cdot f_2 \cdot f_3 \cdot f_4 \cdot f_5$$

$$\cdot\, c_{23;14} \cdot c_{35;14}$$

$$\cdot\, c_{25;134}$$

Example of a five-dimensional R-vine



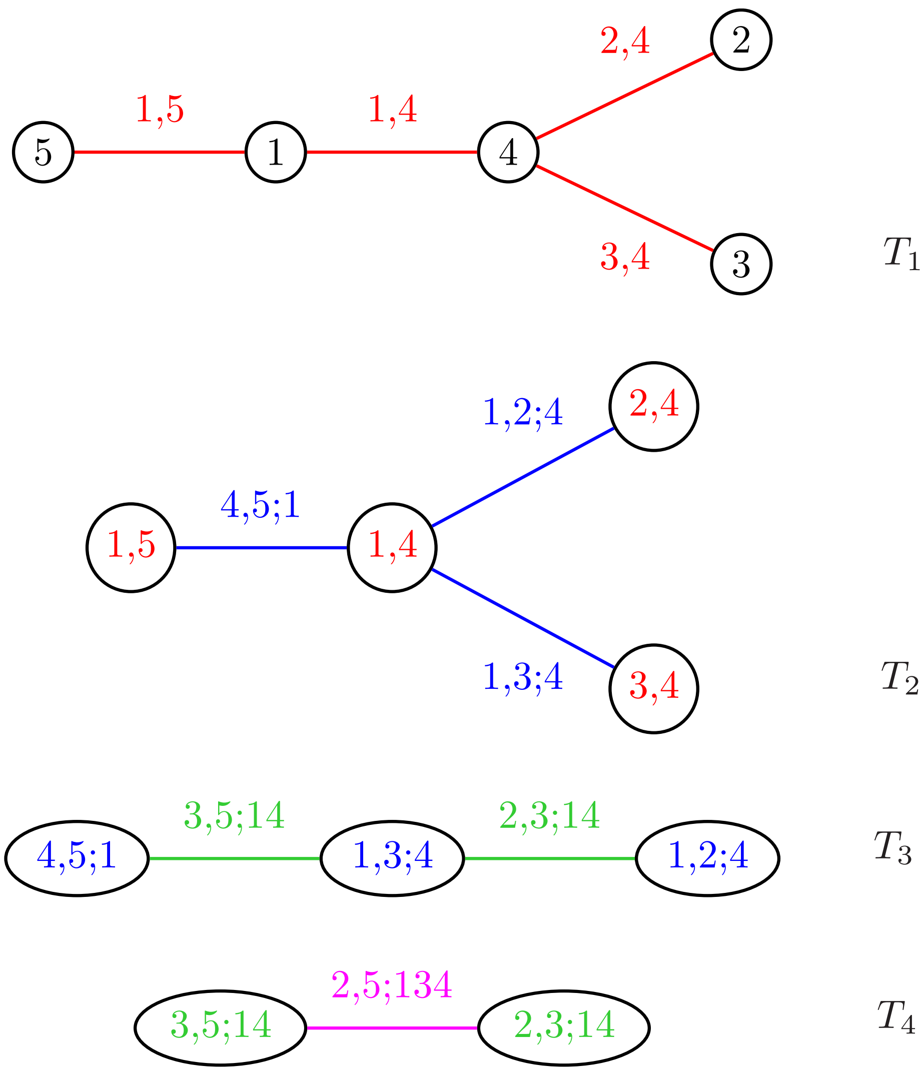
Density

$$f = f_1 \cdot f_2 \cdot f_3 \cdot f_4 \cdot f_5$$

$$\cdot\, c_{23;14} \cdot c_{35;14}$$

$$\cdot\, c_{25;134}$$

Example of a five-dimensional R-vine



Density

$$f = f_1 \cdot f_2 \cdot f_3 \cdot f_4 \cdot f_5$$
$$\cdot\, c_{14} \cdot c_{15} \cdot c_{24} \cdot c_{34}$$
$$\cdot\, c_{12;4} \cdot c_{13;4} \cdot c_{45;1}$$
$$\cdot\, c_{23;14} \cdot c_{35;14}$$
$$\cdot\, c_{25;134}$$

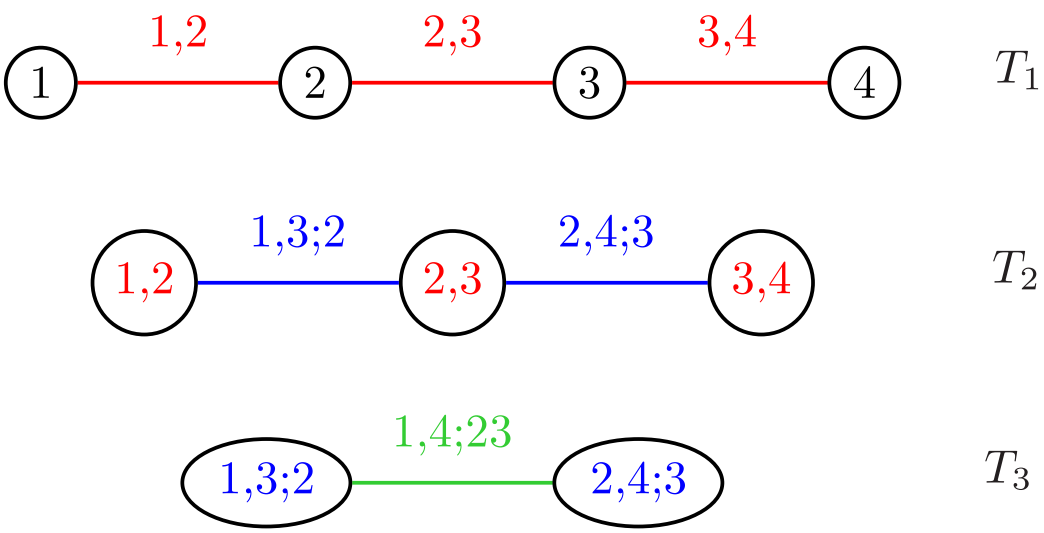
D-vine

An R-vine is called a D-vine if each node in T_1 has a degree of at most 2 (T_1 is a path).

D-vine

An R-vine is called a D-vine if each node in T_1 has a degree of at most 2 (T_1 is a path).

Example ($d = 4$):



Density of D-vine distribution

$$f = f_1 \cdot f_2 \cdot f_3 \cdot f_4 \cdot c_{12} \cdot c_{23} \cdot c_{34} \cdot c_{13;2} \cdot c_{24;3} \cdot c_{14;23}$$

C-vine

An R-vine is called a canonical vine (C-vine) if each tree $T_j, j = 1, ..., d - 1$, has a unique node of degree $d - j$, the *root node*.

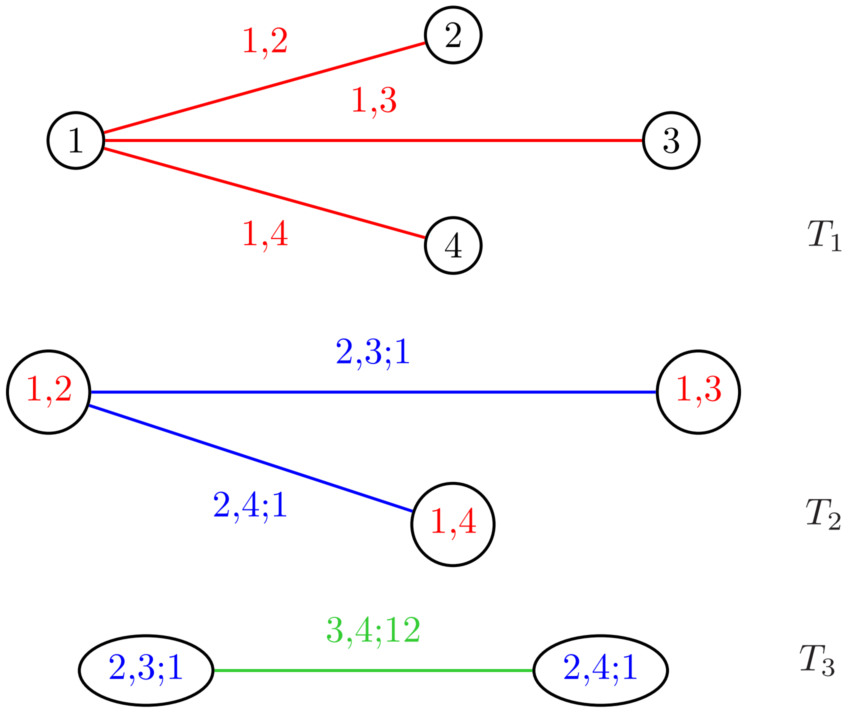
C-vine

An R-vine is called a canonical vine (C-vine) if each tree T_j, $j = 1, ..., d-1$, has a unique node of degree $d-j$, the *root node*.

Density of C-vine distribution

$$
\begin{aligned}
f = {} & f_1 \cdot f_2 \cdot f_3 \cdot f_4 \\
& \cdot c_{12} \cdot c_{13} \cdot c_{14} \\
& \cdot c_{23;1} \cdot c_{24;1} \\
& \cdot c_{34;12}
\end{aligned}
$$

Example ($d = 4$):



Density of an R-vine distribution

For $e \in E_i,\ i \in \{1, ..., d-1\}$, let $e = j(e), k(e)|D(e)$.

- $j(e)$, $k(e)$ = conditioned nodes
- $D(e)$ = conditioning set

This notation is unique for R-vines (see Bedford and Cooke (2002)).

Density of an R-vine distribution

The joint density of an R-vine distribution for $\boldsymbol{X}$ is uniquely determined and given by

$$f(\boldsymbol{x}) = \left[\prod_{k=1}^{d} f_k(x_k)\right] \times \left[\prod_{i=1}^{d-1} \prod_{e \in E_i} c_{j(e),k(e);D(e)}(F_{j|D}(x_{j(e)}|\boldsymbol{x}_{D(e)}), F_{k|D}(x_{k(e)}|\boldsymbol{x}_{D(e)}))\right].$$

$\boldsymbol{x}_{D(e)}$ is the subvector of $\boldsymbol{x} = (x_1, ..., x_d)'$ determined by the indices $D(e)$.

Pair copulas associated with bivariate conditional distributions

Important notation

Let D be an index set not containing i and j.

Pair copulas associated with bivariate conditional distributions

Important notation

Let D be an index set not containing i and j.

- Remember:
  $C_{ij}(u_i, u_j)$ is the copula corresponding to X_i, X_j.

Pair copulas associated with bivariate conditional distributions

Important notation

Let D be an index set not containing i and j.

- Remember:
  $C_{ij}(u_i, u_j)$ is the copula corresponding to X_i, X_j.
- Distinguish:
  $C_{ij;D}(u_i, u_j; \boldsymbol{u}_D)$, the copula corresponding to X_i, X_j given $\boldsymbol{X}_D = \boldsymbol{x}_D$, $\boldsymbol{u}_D = F_D(\boldsymbol{x}_D)$,
  and
  $C_{ij|D}(u_i, u_j | \boldsymbol{u}_D)$, the bivariate density of U_i, U_j given $\boldsymbol{U}_D = \boldsymbol{u}_D$.
  The latter is in general no copula.

Conditional distribution functions I

Let $D = \{j\} \cup D_{-j}$ be an index set with $i \notin D$ and define $\boldsymbol{x}_D = (x_j, \boldsymbol{x}_{D_{-j}})$. Then,

$$f_{i|D}(x_i|\boldsymbol{x}_D) = c_{ij;D_{-j}}(F_{i|D_{-j}}(x_i|\boldsymbol{x}_{D_{-j}}), F_{j|D_{-j}}(x_j|\boldsymbol{x}_{D_{-j}}); \boldsymbol{x}_{D_{-j}}) \cdot f_{i|D_{-j}}(x_i|\boldsymbol{x}_{D_{-j}}).$$

Conditional distribution functions I

Let $D = \{j\} \cup D_{-j}$ be an index set with $i \notin D$ and define $\boldsymbol{x}_D = (x_j, \boldsymbol{x}_{D_{-j}})$. Then,

$$f_{i|D}(x_i|\boldsymbol{x}_D) = c_{ij;D_{-j}}(F_{i|D_{-j}}(x_i|\boldsymbol{x}_{D_{-j}}), F_{j|D_{-j}}(x_j|\boldsymbol{x}_{D_{-j}}); \boldsymbol{x}_{D_{-j}}) \cdot f_{i|D_{-j}}(x_i|\boldsymbol{x}_{D_{-j}}).$$

Univariate case

$$\begin{aligned} F_{i|j}(x_i|x_j) &= \int_{-\infty}^{x_i} f_{i|j}(t|x_j)dt = \int_{-\infty}^{x_i} c_{ij}(F_i(t), F_j(x_j))f_i(t)dt \\ &= \int_{-\infty}^{x_i} \frac{\partial^2 C_{ij}(F_i(t), F_j(x_j))}{\partial F_i(t)\, \partial F_j(x_j)} f_i(t)dt = \frac{\partial\, C_{ij}(F_i(x_i), F_j(x_j))}{\partial F_j(x_j)}. \end{aligned}$$

Conditional distribution functions I

Let $D = \{j\} \cup D_{-j}$ be an index set with $i \notin D$ and define $\boldsymbol{x}_D = (x_j, \boldsymbol{x}_{D_{-j}})$. Then,

$$f_{i|D}(x_i|\boldsymbol{x}_D) = c_{ij;D_{-j}}(F_{i|D_{-j}}(x_i|\boldsymbol{x}_{D_{-j}}), F_{j|D_{-j}}(x_j|\boldsymbol{x}_{D_{-j}}); \boldsymbol{x}_{D_{-j}}) \cdot f_{i|D_{-j}}(x_i|\boldsymbol{x}_{D_{-j}}).$$

Univariate case

$$\begin{aligned} F_{i|j}(x_i|x_j) &= \int_{-\infty}^{x_i} f_{i|j}(t|x_j)dt = \int_{-\infty}^{x_i} c_{ij}(F_i(t), F_j(x_j))f_i(t)dt \\ &= \int_{-\infty}^{x_i} \frac{\partial^2 C_{ij}(F_i(t), F_j(x_j))}{\partial F_i(t)\,\partial F_j(x_j)} f_i(t)dt = \frac{\partial\, C_{ij}(F_i(x_i), F_j(x_j))}{\partial F_j(x_j)}. \end{aligned}$$

- Example:

$$F_{3|2}(x_3|x_2) = \frac{\partial\, C_{23}(F_2(x_2), F_3(x_3))}{\partial F_2(x_2)}$$

Conditional distribution functions I

Let $D = \{j\} \cup D_{-j}$ be an index set with $i \notin D$ and define $\boldsymbol{x}_D = (x_j, \boldsymbol{x}_{D_{-j}})$. Then,

$$f_{i|D}(x_i|\boldsymbol{x}_D) = c_{ij;D_{-j}}(F_{i|D_{-j}}(x_i|\boldsymbol{x}_{D_{-j}}), F_{j|D_{-j}}(x_j|\boldsymbol{x}_{D_{-j}}); \boldsymbol{x}_{D_{-j}}) \cdot f_{i|D_{-j}}(x_i|\boldsymbol{x}_{D_{-j}}).$$

Univariate case

$$\begin{aligned} F_{i|j}(x_i|x_j) &= \int_{-\infty}^{x_i} f_{i|j}(t|x_j)dt = \int_{-\infty}^{x_i} c_{ij}(F_i(t), F_j(x_j))f_i(t)dt \\ &= \int_{-\infty}^{x_i} \frac{\partial^2 C_{ij}(F_i(t), F_j(x_j))}{\partial F_i(t)\, \partial F_j(x_j)} f_i(t)dt = \frac{\partial\, C_{ij}(F_i(x_i), F_j(x_j))}{\partial F_j(x_j)}. \end{aligned}$$

- Example:

$$F_{3|2}(x_3|x_2) = \frac{\partial\, C_{23}(F_2(x_2), F_3(x_3))}{\partial F_2(x_2)}$$

- h-function: Set $h_{ij;D_{-j}}(u_i|u_j; \boldsymbol{u}_{D_{-j}}) := \frac{\partial\, C_{ij;D_{-j}}(u_i,u_j;\boldsymbol{u}_{D_{-j}})}{\partial u_j}$.

  Then $F_{i|j}(x_i|x_j) = h_{ij}(F_i(x_i)|F_j(x_j))$.

Conditional distribution functions II

General case

Under regularity conditions Joe (1996) showed that

$$F_{i|D}(x_i|\boldsymbol{x}_D) = \frac{\partial\, C_{ij;D_{-j}}(F_{i|D_{-j}}(x_i|\boldsymbol{x}_{D_{-j}}), F_{j|D_{-j}}(x_j|\boldsymbol{x}_{D_{-j}}); \boldsymbol{x}_{D_{-j}})}{\partial F_{j|D_{-j}}(x_j|\boldsymbol{x}_{D_{-j}})}.$$

VineCopula: `BiCopHfunc`

Conditional distribution functions II

General case

Under regularity conditions Joe (1996) showed that

$$F_{i|D}(x_i|\boldsymbol{x}_D) = \frac{\partial\, C_{ij;D_{-j}}(F_{i|D_{-j}}(x_i|\boldsymbol{x}_{D_{-j}}), F_{j|D_{-j}}(x_j|\boldsymbol{x}_{D_{-j}}); \boldsymbol{x}_{D_{-j}})}{\partial F_{j|D_{-j}}(x_j|\boldsymbol{x}_{D_{-j}})}.$$

- $F_{i|D}(x_i|\boldsymbol{x}_D) = h_{ij|D_{-j}}(F_{i|D_{-j}}(x_i|\boldsymbol{x}_{D_{-j}}), F_{j|D_{-j}}(x_j|\boldsymbol{x}_{D_{-j}}); \boldsymbol{x}_{D_{-j}})$
  $\rightarrow$ recursive computation!

VineCopula: `BiCopHfunc`

Conditional distribution functions II

General case

Under regularity conditions Joe (1996) showed that

$$F_{i|D}(x_i|\boldsymbol{x}_D) = \frac{\partial\, C_{ij;D_{-j}}(F_{i|D_{-j}}(x_i|\boldsymbol{x}_{D_{-j}}), F_{j|D_{-j}}(x_j|\boldsymbol{x}_{D_{-j}}); \boldsymbol{x}_{D_{-j}})}{\partial F_{j|D_{-j}}(x_j|\boldsymbol{x}_{D_{-j}})}.$$

- $F_{i|D}(x_i|\boldsymbol{x}_D) = h_{ij|D_{-j}}(F_{i|D_{-j}}(x_i|\boldsymbol{x}_{D_{-j}}), F_{j|D_{-j}}(x_j|\boldsymbol{x}_{D_{-j}}); \boldsymbol{x}_{D_{-j}})$
  $\rightarrow$ recursive computation!
- Example:

$$\begin{aligned} F_{3|12}(x_3|x_1, x_2) &= \frac{\partial\, C_{13;2}(F_{1|2}(x_1|x_2), F_{3|2}(x_3|x_2); x_2)}{\partial F_{1|2}(x_1|x_2)} \\ &= h_{13;2}(F_{1|2}(x_1|x_2)|F_{3|2}(x_3|x_2); x_2) \\ &= h_{13;2}\big(h_{12}(F_1(x_1)|F_2(x_2))\big|h_{23}(F_3(x_3)|F_2(x_2)); x_2\big) \end{aligned}$$

VineCopula: `BiCopHfunc`

Simplifying assumption

- To facilitate inference of vine copulas, pair-copulas are chosen to be independent of conditioning values. Arguments however depend on the conditioning values.

$$c_{j,k;D}(F_{j|D}(x_j|\boldsymbol{x}_D), F_{k|D}(x_k|\boldsymbol{x}_D); \boldsymbol{x}_D) \equiv c_{j,k;D}(F_{j|D}(x_j|\boldsymbol{x}_D), F_{k|D}(x_k|\boldsymbol{x}_D))$$

- Hobæk Haff et al. (2010) and Stöber et al. (2013) give examples where the pair-copula parameters depend on the specific conditioning values. Recent and ongoing investigation in Acar et al. (2012) and Killiches et al. (2016).

- Hobæk Haff et al. (2010) show that this restriction is not severe in examples.

Simplifying assumption

Copulas for which the simplifying assumption is fulfilled:

- multivariate Gaussian copula
- multivariate Student's t copula (only one arising from scale mixtures of normals, see Stöber et al. (2013))
- partial correlations $\rho_{ij;D}$ are copula parameters in a Gaussian or Student's t-vine with common degree of freedom; degree-of-freedom increase by 1 as tree number increase by 1
- multivariate Clayton copula (the only Archimedean; Takahashi (1965), Stöber et al. (2013))

Some more remarks

- Number of different R-vines is huge (Morales-Nápoles 2011).
- Flexibility is added by allowing for different pair-copula families.

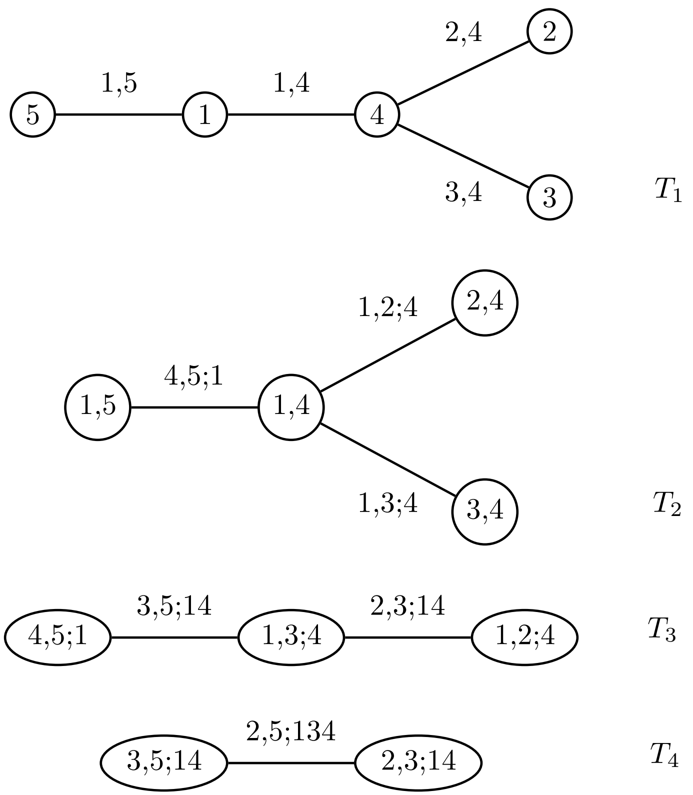
Tractable estimation and model selection methods are vital.

R-vine structure matrices

Efficient encoding of R-vine models needed for statistical inference.

- Matrix notation by Morales-Nápoles et al. (2010) and Dißmann et al. (2013).

$$\begin{pmatrix} 2 & & & & \\ 5 & 3 & & & \\ 3 & 5 & 4 & & \\ 1 & 1 & 5 & 5 & \\ 4 & 4 & 1 & 1 & 1 \end{pmatrix}$$



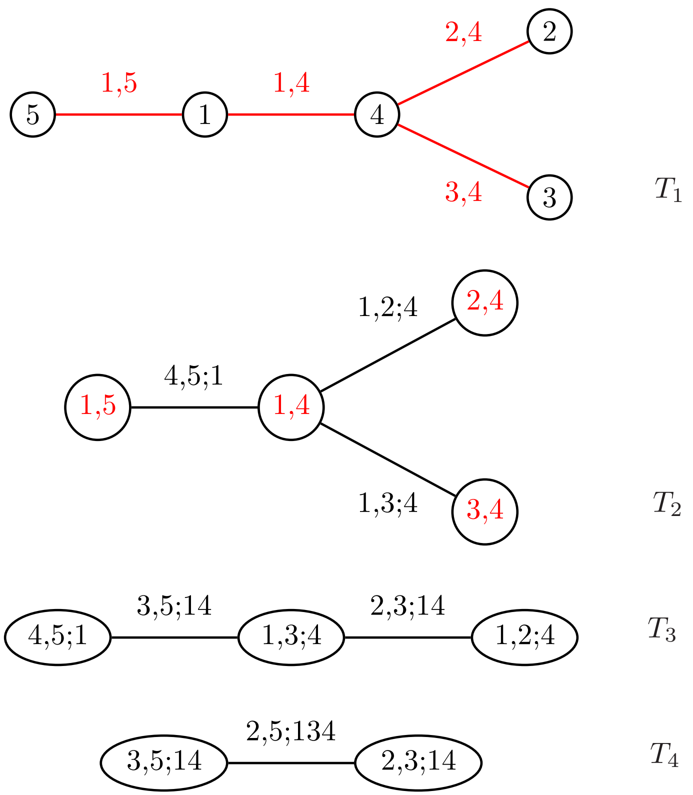
R-vine structure matrices

Efficient encoding of R-vine models needed for statistical inference.

- Matrix notation by Morales-Nápoles et al. (2010) and Dißmann et al. (2013).

$$\begin{pmatrix} 2 & & & & \\ 5 & 3 & & & \\ 3 & 5 & 4 & & \\ 1 & 1 & 5 & 5 & \\ 4 & 4 & 1 & 1 & 1 \end{pmatrix}$$

1. $\{2,4\},\{3,4\},\{4,1\},\{5,1\}$



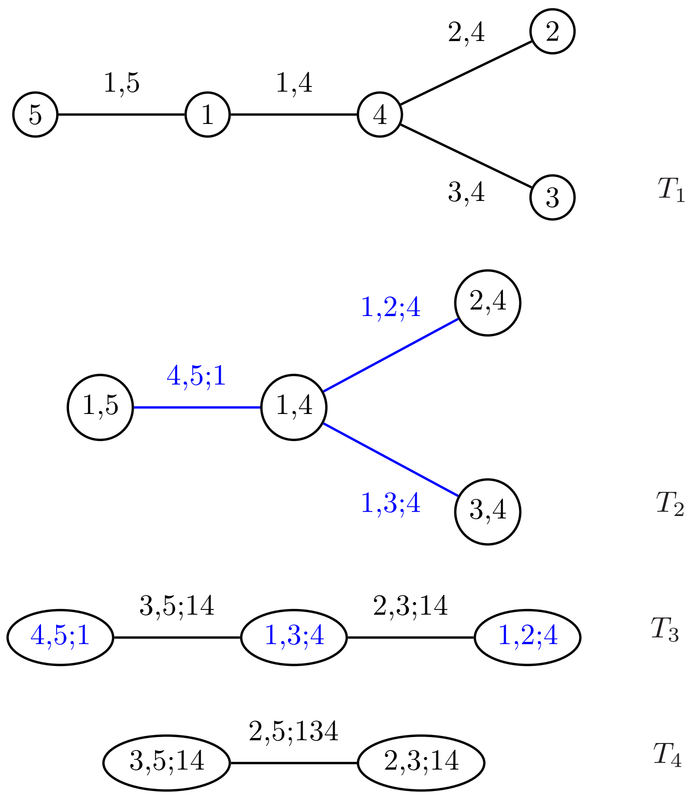
R-vine structure matrices

Efficient encoding of R-vine models needed for statistical inference.

- Matrix notation by Morales-Nápoles et al. (2010) and Dißmann et al. (2013).

$$\begin{pmatrix} 2 & & & & \\ 5 & 3 & & & \\ 3 & 5 & 4 & & \\ 1 & 1 & 5 & 5 & \\ \hline 4 & 4 & 1 & 1 & 1 \end{pmatrix}$$

1. $\{2,4\}, \{3,4\}, \{4,1\}, \{5,1\}$
2. $\{2,1;4\}, \{3,1;4\}, \{4,5;1\}$



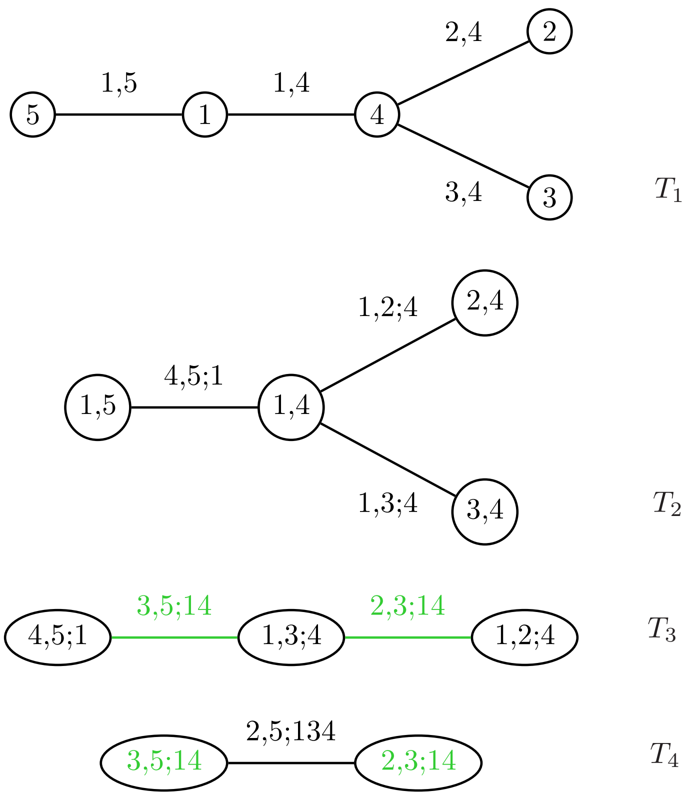
R-vine structure matrices

Efficient encoding of R-vine models needed for statistical inference.

- Matrix notation by Morales-Nápoles et al. (2010) and Dißmann et al. (2013).

$$\begin{pmatrix} 2 & & & & \\ 5 & 3 & & & \\ 3 & 5 & 4 & & \\ 1 & 1 & 5 & 5 & \\ 4 & 4 & 1 & 1 & 1 \end{pmatrix}$$

1. $\{2,4\},\{3,4\},\{4,1\},\{5,1\}$
2. $\{2,1;4\},\{3,1;4\},\{4,5;1\}$
3. $\{2,3;14\},\{3,5;14\}$



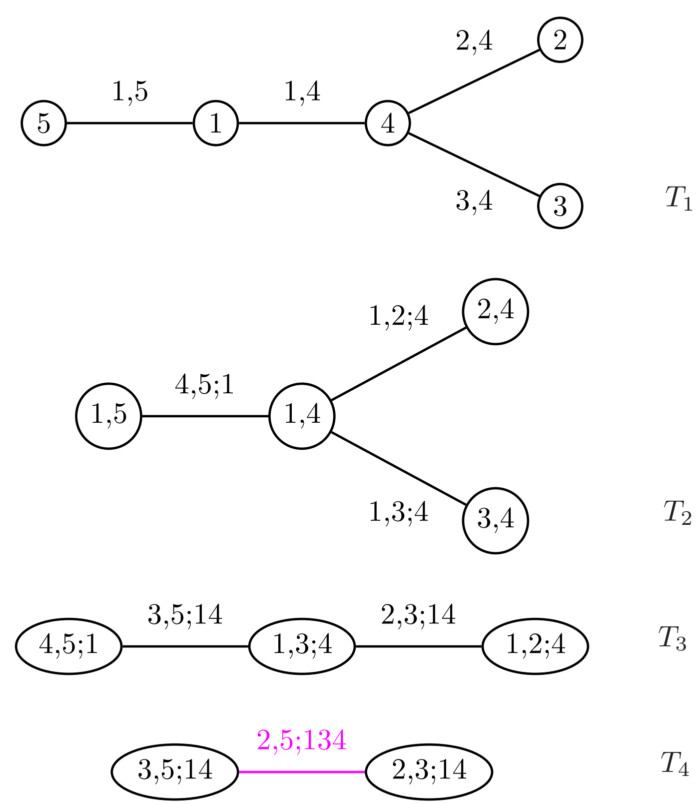
R-vine structure matrices

Efficient encoding of R-vine models needed for statistical inference.

- Matrix notation by Morales-Nápoles et al. (2010) and Dißmann et al. (2013).

$$\begin{pmatrix} 2 & & & & \\ 5 & 3 & & & \\ \hline 3 & 5 & 4 & & \\ 1 & 1 & 5 & 5 & \\ 4 & 4 & 1 & 1 & 1 \end{pmatrix}$$

1. $\{2,4\}, \{3,4\}, \{4,1\}, \{5,1\}$
2. $\{2,1;4\}, \{3,1;4\}, \{4,5;1\}$
3. $\{2,3;14\}, \{3,5;14\}$
4. $\{2,5;314\}$



R-vine copula and parameter matrices

Copula families and parameters can be stored in associated matrices.

$$\begin{pmatrix} 2 & & & & \\ 5 & 3 & & & \\ 3 & 5 & 4 & & \\ 1 & 1 & 5 & 5 & \\ 4 & 4 & 1 & 1 & 1 \end{pmatrix} \longrightarrow \begin{pmatrix} C_{25;314} & & & & \\ C_{23;14} & C_{35;14} & & & \\ C_{21;4} & C_{31;4} & C_{45;1} & & \\ C_{24} & C_{34} & C_{41} & C_{51} & \end{pmatrix}$$

$$\boldsymbol{\theta} = \begin{pmatrix} \theta_{25;314} & & & & \\ \theta_{23;14} & \theta_{35;14} & & & \\ \theta_{21;4} & \theta_{31;4} & \theta_{45;1} & & \\ \theta_{24} & \theta_{34} & \theta_{41} & \theta_{51} & \end{pmatrix}$$

R-vine matrix objects

An `RVineMatrix` object contains all required matrices:

```
> Matrix = matrix(c(2,0,0,0,0,
+                   5,3,0,0,0,
+                   4,5,4,0,0,
+                   1,1,5,5,0,
+                   4,4,1,1,1),5,5)
> family = matrix(c(0,0,0,0,0,
+                   1,0,0,0,0,
+                   3,3,0,0,0,
+                   4,4,4,0,0,
+                   4,1,1,3,0),5,5)
> par = matrix(c(0  ,0  ,0  ,0  ,0,
+                0.2,0  ,0  ,0  ,0,
+                0.9,1.1,0  ,0  ,0,
+                1.5,1.6,1.9,0  ,0,
+                3.9,0.9,0.5,4.ol8,0),5,5)
> RVM = RVineMatrix(Matrix=Matrix, family=family, par=par,
+      par2=matrix(0,5,5), names=c("V1","V2","V3","V4","V5"))
```

Summary

Vine copulas $(\mathcal{V}, \boldsymbol{B}, \boldsymbol{\theta})$ have three components:

- structure $\mathcal{V}$,
- pair-copulas $\boldsymbol{B} = \boldsymbol{B}(\mathcal{V})$ and
- parameters $\boldsymbol{\theta} = \boldsymbol{\theta}(\boldsymbol{B}(\mathcal{V}))$.

Relations between model components have to be respected in inference!

Lecture V

Estimation and model selection for vine copulas

Conditional inverse method

Aim: Sample from multivariate distribution F.

1. Obtain d i.i.d. uniform samples $(v_1, ..., v_d)$.
2. Set

$$
\begin{aligned}
x_1 &:= F_1^{-1}(v_1) \\
x_2 &:= F_{2|1}^{-1}(v_2|x_1) \\
x_3 &:= F_{3|12}^{-1}(v_3|x_1, x_2) \\
&\vdots \\
x_d &:= F_{d|1,\dots,d-1}^{-1}(v_d|x_1, \dots, x_{d-1}).
\end{aligned}
$$

3. Then $\boldsymbol{x} := (x_1, ..., x_d)'$ is a sample from F (see, e.g., Devroye (1986)).

Conditional inverse method

Aim: Sample from multivariate distribution F.

1. Obtain d i.i.d. uniform samples $(v_1, ..., v_d)$.
2. Set

$$\begin{aligned} x_1 &:= F_1^{-1}(v_1) \\ x_2 &:= F_{2|1}^{-1}(v_2|x_1) \\ x_3 &:= F_{3|12}^{-1}(v_3|x_1, x_2) \\ &\vdots \\ x_d &:= F_{d|1,...,d-1}^{-1}(v_d|x_1, ..., x_{d-1}). \end{aligned}$$

3. Then $\boldsymbol{x} := (x_1, ..., x_d)'$ is a sample from F (see, e.g., Devroye (1986)).

Note: Let C be the copula associated to F, then it is sufficient to obtain a sample $\boldsymbol{u} := (u_1, ..., u_d)'$ from C and set $x_j := F_j^{-1}(u_j),\ j = 1, ..., d$.

Simulating vine copulas

Question: How do inverse conditional distribution functions look like for vine copulas?

Simulating vine copulas

Question: How do inverse conditional distribution functions look like for vine copulas?

Example ($d = 3$):

1. Obtain 3 i.i.d. uniform samples (v_1, v_2, v_3).

General sampling algorithm in Dißmann et al. (2013).

VineCopula: `RVineSim`

Simulating vine copulas

Question: How do inverse conditional distribution functions look like for vine copulas?

Example ($d = 3$):

1. Obtain 3 i.i.d. uniform samples (v_1, v_2, v_3).
2. Set

$$u_1 := v_1$$

$$= h_{23}^{-1}\left(h_{13;2}^{-1}\big(v_3 \big| h_{12}(u_1|u_2); u_2\big) \big| u_2\right).$$

Recall: $F_{3|12}(u_3|u_1, u_2) = h_{13;2}\big(h_{23}(u_3|u_2) \big| h_{12}(u_1|u_2); u_2\big)$

General sampling algorithm in Dißmann et al. (2013).

VineCopula: `RVineSim`

Simulating vine copulas

Question: How do inverse conditional distribution functions look like for vine copulas?

Example ($d = 3$):

1. Obtain 3 i.i.d. uniform samples (v_1, v_2, v_3).
2. Set

$$u_1 := v_1$$

$$= h_{23}^{-1}\left(h_{13;2}^{-1}\big(v_3 | h_{12}(u_1 | u_2); u_2\big) | u_2\right).$$

3. Then $\boldsymbol{u} := (u_1, u_2, u_3)'$ is a sample from the vine copula.

General sampling algorithm in Dißmann et al. (2013).

VineCopula: `RVineSim`

Simulating vine copulas

Question: How do inverse conditional distribution functions look like for vine copulas?

Example ($d = 3$):

1. Obtain 3 i.i.d. uniform samples (v_1, v_2, v_3).
2. Set

$$\begin{aligned}
u_1 &:= v_1 \\
u_2 &:= F_{2|1}^{-1}(v_2|u_1) = h_{12}^{-1}(v_2|u_1) \\
u_3 &:= F_{3|12}^{-1}(v_3|u_1, u_2) = h_{23}^{-1}\left(h_{13;2}^{-1}\big(v_3|h_{12}(u_1|u_2); u_2\big)|u_2\right).
\end{aligned}$$

Recall: $F_{3|12}(u_3|u_1, u_2) = h_{13;2}\Big(h_{23}(u_3|u_2)\Big|h_{12}(u_1|u_2); u_2\Big)$

Simulating vine copulas

Question: How do inverse conditional distribution functions look like for vine copulas?

Example ($d = 3$):

1. Obtain 3 i.i.d. uniform samples (v_1, v_2, v_3).
2. Set

$$
\begin{aligned}
u_1 &:= v_1 \\
u_2 &:= F_{2|1}^{-1}(v_2|u_1) = h_{12}^{-1}(v_2|u_1) \\
u_3 &:= F_{3|12}^{-1}(v_3|u_1, u_2) = h_{23}^{-1}\left(h_{13;2}^{-1}\big(v_3|h_{12}(u_1|u_2); u_2\big)\Big|u_2\right).
\end{aligned}
$$

Recall: $F_{3|12}(u_3|u_1, u_2) = h_{13;2}\Big(h_{23}(u_3|u_2)\Big|h_{12}(u_1|u_2); u_2\Big)$

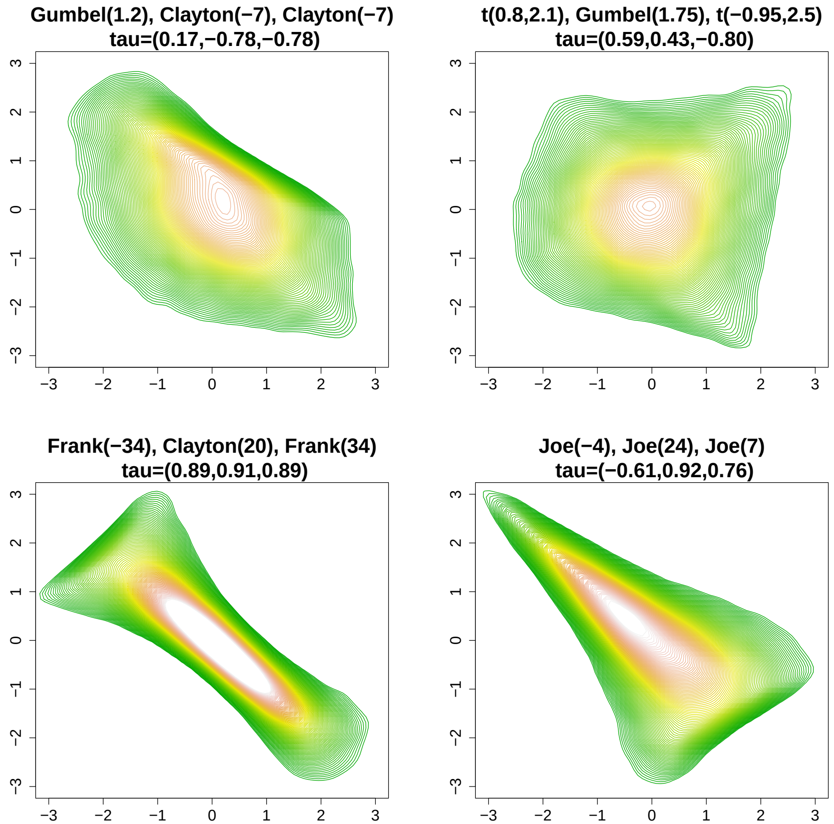
3. Then $\boldsymbol{u} := (u_1, u_2, u_3)'$ is a sample from the vine copula.

General sampling algorithm in Dißmann et al. (2013).

VineCopula: `RVineSim`

Illustration of 3-dimensional D-vine

Pairs {1,2} and {2,3} are modeled unconditionally in tree 1. Contours of bivariate {1,3} margin with standard normal margins after integration:



Simulation

Remember, we defined an `RVM-Object` with an R-vine-structure, pair-copula families and their parameters and stored it in `RVM`.

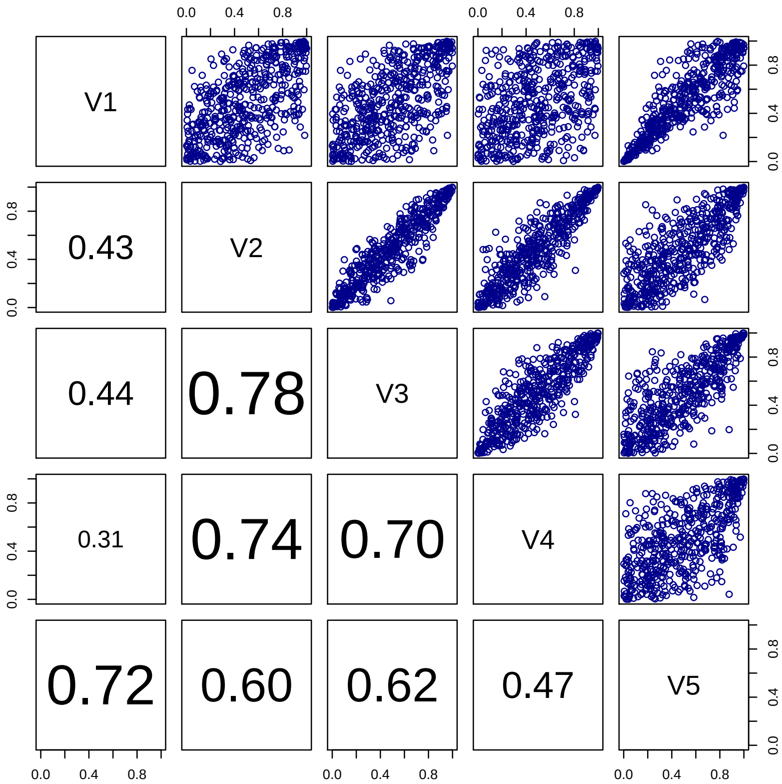
```
> simdat = RVineSim(500, RVM)
```

Simulation

Remember, we defined an `RVM-Object` with an R-vine-structure, pair-copula families and their parameters and stored it in `RVM`.

```
> simdat = RVineSim(500, RVM)

> head(simdat)
       V1   V2   V3   V4   V5
[1,] 0.51 0.24 0.42 0.33 0.45
[2,] 0.23 0.14 0.16 0.12 0.20
[3,] 0.65 0.38 0.46 0.29 0.70
[4,] 0.43 0.18 0.08 0.08 0.26
[5,] 0.86 0.86 0.85 0.86 0.87
[6,] 0.71 0.71 0.80 0.68 0.88
```



General remarks on copula estimation

Marginal distributions F_{γ_j} and copula C_θ have to be estimated.

- Joint maximum likelihood (ML) estimation

$$\begin{pmatrix}\widehat{\theta}_{\text{ML}}\\ \widehat{\gamma}_{\text{ML}}\end{pmatrix} = \operatorname*{argmax}_{\theta,\gamma} \sum_{i=1}^{n} \left(\log\left[c_\theta(F_{\gamma_1}(x_{i1}), \ldots, F_{\gamma_d}(x_{id}))\right] + \sum_{j=1}^{d} \log f_{\gamma_j}(x_{ij}) \right)$$

- Inference functions for margins (IFM) (Joe and Xu 1996)

$$\widehat{\theta}_{\text{IFM}} = \operatorname*{argmax}_{\theta} \sum_{i=1}^{n} \log\left[c_\theta(F_{\hat{\gamma}_1}(x_{i1}), \ldots, F_{\hat{\gamma}_d}(x_{id}))\right]$$

- Maximum pseudo likelihood (MPL) estimation (Genest et al. 1995)

$$\widehat{\theta}_{\text{MPL}} = \operatorname*{argmax}_{\theta} \sum_{i=1}^{n} \log\left[c_\theta\left(u_{i1}, \ldots, u_{id}\right)\right],$$

  where $u_{ij} = r_{ij}/(n+1)$ are transformed ranks.

We assume data is already marginally uniform ($\rightarrow$ IFM, MPL).

Sequential estimation

Parameters are sequentially estimated starting from the top tree.

Sequential estimation

Parameters are sequentially estimated starting from the top tree.

Example ($d = 3$):

- Parameters: $\boldsymbol{\theta} = (\theta_{12}, \theta_{23}, \theta_{13|2})'$
- Observations: $\{(x_{i1}, x_{i2}, x_{i3}), i = 1, ..., n\}$

Sequential estimation

Parameters are sequentially estimated starting from the top tree.

Example ($d = 3$):

- Parameters: $\boldsymbol{\theta} = (\theta_{12}, \theta_{23}, \theta_{13|2})'$
- Observations: $\{(x_{i1}, x_{i2}, x_{i3}), i = 1, \dots, n\}$

1. Tree 1:
   - Estimate θ_{12} from $\{(x_{i1}, x_{i2}), i = 1, \dots, n\}$.
   - Estimate θ_{23} from $\{(x_{i2}, x_{i3}), i = 1, \dots, n\}$.

Sequential estimation

Parameters are sequentially estimated starting from the top tree.

Example ($d = 3$):

- Parameters: $\boldsymbol{\theta} = (\theta_{12}, \theta_{23}, \theta_{13|2})'$
- Observations: $\{(x_{i1}, x_{i2}, x_{i3}), i = 1, ..., n\}$

1. Tree 1:
   - Estimate θ_{12} from $\{(x_{i1}, x_{i2}), i = 1, ..., n\}$.
   - Estimate θ_{23} from $\{(x_{i2}, x_{i3}), i = 1, ..., n\}$.
2. Tree 2:
   - Define pseudo observations
   $$\hat{v}_{i,1|2} := F_{1|2}(x_{i1}|x_{i2}; \hat{\theta}_{12}) \text{ and } \hat{v}_{i,3|2} := F_{3|2}(x_{i3}|x_{i2}; \hat{\theta}_{23}).$$
   - Estimate $\theta_{13;2}$ from $\{(\hat{v}_{i,1|2}, \hat{v}_{i,3|2}), i = 1, ..., n\}$.

Sequential estimation

Parameters are sequentially estimated starting from the top tree.

Example ($d = 3$):

- Parameters: $\boldsymbol{\theta} = (\theta_{12}, \theta_{23}, \theta_{13|2})'$
- Observations: $\{(x_{i1}, x_{i2}, x_{i3}), i = 1, ..., n\}$

1. Tree 1:
   - Estimate θ_{12} from $\{(x_{i1}, x_{i2}), i = 1, ..., n\}$.
   - Estimate θ_{23} from $\{(x_{i2}, x_{i3}), i = 1, ..., n\}$.
2. Tree 2:
   - Define pseudo observations

$$\hat{v}_{i,1|2} := F_{1|2}(x_{i1}|x_{i2}; \hat{\theta}_{12}) \text{ and } \hat{v}_{i,3|2} := F_{3|2}(x_{i3}|x_{i2}; \hat{\theta}_{23}).$$

   - Estimate $\theta_{13;2}$ from $\{(\hat{v}_{i,1|2}, \hat{v}_{i,3|2}), i = 1, ..., n\}$.

Asymptotic theory is available (Hobæk Haff 2013), however analytical standard errors are difficult to compute.

Maximum likelihood estimation

Example ($d = 3$):

$$\widehat{\boldsymbol{\theta}} = \underset{\boldsymbol{\theta}}{\operatorname{argmax}} \sum_{i=1}^{n} \Big(\log c_{12}(F_1(x_{i1}), F_2(x_{i2}); \theta_{12}) + \log c_{23}(F_2(x_{i2}), F_3(x_{i3}); \theta_{23})$$

$$+ \log c_{13;2}(F_{1|2}(x_{i1}|x_{i2}; \theta_{12}), F_{3|2}(x_{i3}|x_{i2}; \theta_{23}); \theta_{13;2}) \Big)$$

Maximum likelihood estimation

Example ($d = 3$):

$$\widehat{\boldsymbol{\theta}} = \underset{\boldsymbol{\theta}}{\operatorname{argmax}} \sum_{i=1}^{n} \Big(\log c_{12}(F_1(x_{i1}), F_2(x_{i2}); \theta_{12}) + \log c_{23}(F_2(x_{i2}), F_3(x_{i3}); \theta_{23}) \\ + \log c_{13;2}(F_{1|2}(x_{i1}|x_{i2}; \theta_{12}), F_{3|2}(x_{i3}|x_{i2}; \theta_{23}); \theta_{13;2}) \Big)$$

- Asymptotically efficient under regularity conditions.
- Estimates of standard errors can be based on inverse Hessian matrix (Stöber and Schepsmeier 2013).
- Sequential estimates can be used as starting values.
- Numerical problems for large dimensions, i.e. negative variance estimates might occur.

Likelihood computation for R-vine copulas

Sequential and maximum likelihood estimation look simple

but identification of required conditional distribution functions not trivial.

Likelihood computation for R-vine copulas

Sequential and maximum likelihood estimation look simple
but identification of required conditional distribution functions not trivial.

Example ($d = 5$): Evaluate $c_{2,5;134}$.

$$c = c_{14} \cdot c_{15} \cdot c_{24} \cdot c_{34} \cdot c_{12;4} \cdot c_{13;4} \cdot c_{45;1} \cdot c_{23;14} \cdot c_{35;14} \cdot c_{25;134}$$

$$c_{3,5;14} \qquad\qquad c_{2,3;14}$$

$$F_{3|145} \quad F_{5|134} \qquad F_{2|134} \quad F_{3|124}$$

Likelihood computation for R-vine copulas

Sequential and maximum likelihood estimation look simple
but identification of required conditional distribution functions not trivial.

Example ($d = 5$): Evaluate $c_{2,5;134}$.

$$c = c_{14} \cdot c_{15} \cdot c_{24} \cdot c_{34} \cdot c_{12;4} \cdot c_{13;4} \cdot c_{45;1} \cdot c_{23;14} \cdot c_{35;14} \cdot c_{25;134}$$

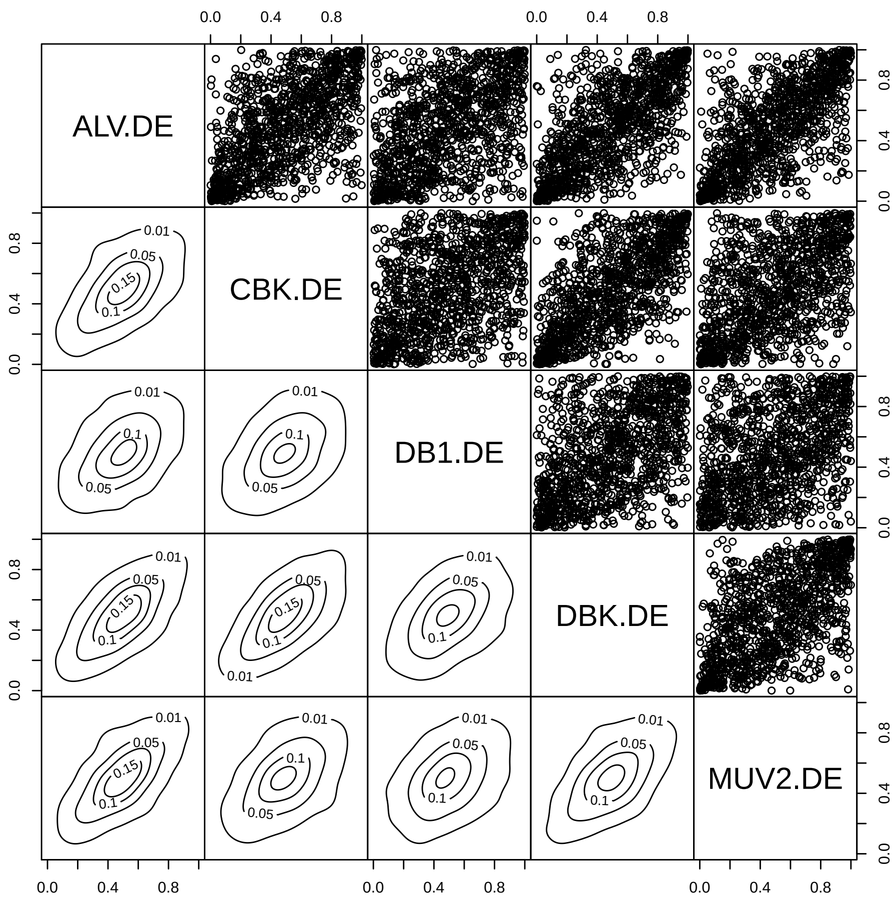
$$\begin{tikzcd}
 & c_{3,5;14} \arrow[dl] \arrow[dr] & & & & c_{2,3;14} \arrow[dl] \arrow[dr] & \\
\cancel{F_{3|145}} & & F_{5|134} \arrow[dr] & & F_{2|134} \arrow[dl] & & \cancel{F_{3|124}} \\
 & & & c_{2,5;134} & & &
\end{tikzcd}$$

- Dißmann et al. (2013) give algorithm to evaluate an R-vine density.

Example

- Daily log returns of 5 major German stocks.
  - Deutsche Bank (`DBK.DE`)
  - Commerzbank (`CBK.DE`)
  - Allianz (`ALV.DE`)
  - Munich Re (`MUV2.DE`)
  - Deutsche Börse (`DB1.DE`)
- Observed from January 2005 to August 2009 (1158 observations).
- Time series are filtered using GARCH(1,1) with Student's t innovations.
- Data set of standardized residuals transformed to [0,1].

A first look at the data



Parameter estimation I

- Sequential estimation (based on `BiCopEst`)
  - either using bivariate inversion of Kendall's τ:

    ```
    > RVineSeqEst(data, RVM, method="itau")
    ```
  - or bivariate maximum likelihood estimation:

    ```
    > RVineSeqEst(data, RVM, method="mle")
    ```

- Very fast, since only bivariate estimation.
- Provides good starting values for joint maximum likelihood estimation.

Parameter estimation II

- Maximum likelihood estimation of all parameters jointly (log-likelihood computation: `RVineLogLik`).

```
> RVineMLE(data, RVM, start, start2, maxit,
+          grad, hessian, se)
```

Parameter estimation II

- Maximum likelihood estimation of all parameters jointly (log-likelihood computation: `RVineLogLik`).

```
> RVineMLE(data, RVM, start=0, start2=0, maxit,
+          grad, hessian, se)
```

▶ Starting values can be calculated using sequential estimation.

Parameter estimation II

- Maximum likelihood estimation of all parameters jointly (log-likelihood computation: `RVineLogLik`).

```
> RVineMLE(data, RVM, start, start2, maxit,
+          grad=TRUE, hessian, se)
```

- Starting values can be calculated using sequential estimation.
- Analytical gradient can be used for numerical optimization (see `RVineGrad`).

Parameter estimation II

- Maximum likelihood estimation of all parameters jointly (log-likelihood computation: `RVineLogLik`).

```
> RVineMLE(data, RVM, start, start2, maxit,
+          grad, hessian=TRUE, se=TRUE)
```

- Starting values can be calculated using sequential estimation.
- Analytical gradient can be used for numerical optimization (see `RVineGrad`).
- Standard errors can be computed based on the analytical Hessian (see `RVineStdError` and `RVineHessian`).
- In `RVineMLE(...,hessian=TRUE)` a numerical Hessian is returned.

Parameter estimation III

```
> mle = RVineMLE(data=dax, RVM, start=0, start2=0,
+                grad=TRUE, hessian=TRUE, se=TRUE)
```

Parameter estimation III

```
> mle = RVineMLE(data=dax, RVM, start=0, start2=0,
+                 grad=TRUE, hessian=TRUE, se=TRUE)
```

```
> mle$RVM$par
     [,1] [,2] [,3] [,4] [,5]
[1,] 0.00 0.00 0.00 0.00    0
[2,] 0.01 0.00 0.00 0.00    0
[3,] 0.19 0.08 0.00 0.00    0
[4,] 1.13 1.10 1.12 0.00    0
[5,] 1.89 0.50 0.71 1.47    0
```

Parameter estimation III

```
> mle = RVineMLE(data=dax, RVM, start=0, start2=0,
+                  grad=TRUE, hessian=TRUE, se=TRUE)
```

```
> mle$RVM$par
     [,1] [,2] [,3] [,4] [,5]
[1,] 0.00 0.00 0.00 0.00    0
[2,] 0.01 0.00 0.00 0.00    0
[3,] 0.19 0.08 0.00 0.00    0
[4,] 1.13 1.10 1.12 0.00    0
[5,] 1.89 0.50 0.71 1.47    0
```

```
> mle$se
     [,1] [,2] [,3] [,4] [,5]
[1,] 0.00 0.00 0.00 0.00    0
[2,] 0.03 0.00 0.00 0.00    0
[3,] 0.05 0.03 0.00 0.00    0
[4,] 0.02 0.02 0.03 0.00    0
[5,] 0.05 0.02 0.01 0.07    0
```

- Standard errors for Kendall's τ can be estimated using the delta method.
- `RVineMLE` returns std. errors based on the numerical Hessian matrix
- An analytical Hessian matrix can be calculated by `RVineHessian`

Pair-copula selection

Problematic because of small Kullback-Leibler probability distances and boundary cases (Student's t, two parameter Archimedean: Clayton-Gumbel (BB1), Joe-Clayton (BB7),...).

Pair-copula selection

Problematic because of small Kullback-Leibler probability distances and boundary cases (Student's t, two parameter Archimedean: Clayton-Gumbel (BB1), Joe-Clayton (BB7),...).

Many different approaches available:

- Graphical tools: scatter plots, empirical contour plots,...
- Copula goodness-of-fit tests: Choose family with highest *p*-value (if larger than α).
- Choose family with highest likelihood or smallest AIC/BIC/....
- Choose family which best reproduces data characteristics (Kendall's τ, joint tail behavior).
- ...

Pair-copula selection

Independence test (Genest and Favre 2007)

The test exploits the approximate standard normality of the test statistic

$$\texttt{statistic} := T = \sqrt{\frac{9N(N-1)}{2(2N+5)}} \times |\hat{\tau}|,$$

where N, the number of observations, is large and $\hat{\tau}$ is the empirical Kendall's tau of the data vectors u_1 and u_2. The p-value of the null hypothesis of bivariate independence hence is asymptotically

$$\texttt{p.value} = 2 \times (1 - \Phi(T)),$$

where Φ is the standard normal distribution function.

Pair copula selection

```
> cops = RVineCopSelect(data=dax, familyset=NA,
+                        Matrix=Matrix, selectioncrit="AIC",
+                        indeptest=FALSE, level=0.05)
```

`RVineCopSelect` uses the sequential estimation approach to estimate the necessary copula parameters.

Pair copula selection

```
> cops = RVineCopSelect(data=dax, familyset=NA,
+                        Matrix=Matrix, selectioncrit="AIC",
+                        indeptest=FALSE, level=0.05)
```

`RVineCopSelect` uses the sequential estimation approach to estimate the necessary copula parameters.

```
> cops$family
     [,1] [,2] [,3] [,4] [,5]
[1,]    0    0    0    0    0
[2,]    4    0    0    0    0
[3,]    5   14    0    0    0
[4,]    2    2    2    0    0
[5,]    2   20    2    2    0
```

Treewise construction of R-vines (Tree 1)

Dißmann et al. (2013): Capture strong dependencies of data x_{ij}, $j = 1, ..., d$, $i = 1, ..., n$..

1. Calculate an empirical dependence measure $\hat{\delta}_{jk}$ for all possible variable pairs $\{jk\}$ ($\rightarrow$ edge weights).

Treewise construction of R-vines (Tree 1)

Dißmann et al. (2013): Capture strong dependencies of data x_{ij}, $j = 1, \dots, d,\ i = 1, \dots, n$..

1. Calculate an empirical dependence measure $\hat{\delta}_{jk}$ for all possible variable pairs $\{jk\}$ ($\rightarrow$ edge weights).
2. Select the tree on all nodes that maximizes the sum of absolute empirical dependencies ($\rightarrow$ maximum spanning tree):

$$\max \sum_{\substack{\text{edges } e=\{j,k\} \text{ in} \\ \text{spanning tree}}} |\hat{\delta}_{jk}|.$$

Treewise construction of R-vines (Tree 1)

Dißmann et al. (2013): Capture strong dependencies of data x_{ij}, $j = 1, \ldots, d,\ i = 1, \ldots, n$..

1. Calculate an empirical dependence measure $\hat{\delta}_{jk}$ for all possible variable pairs $\{jk\}$ ($\rightarrow$ edge weights).
2. Select the tree on all nodes that maximizes the sum of absolute empirical dependencies ($\rightarrow$ maximum spanning tree):

$$\max \sum_{\substack{\text{edges } e=\{j,k\} \text{ in} \\ \text{spanning tree}}} |\hat{\delta}_{jk}|.$$

Parsimonious model selection: choose independence copula if possible.

Treewise construction of R-vines (Tree 1)

Dißmann et al. (2013): Capture strong dependencies of data x_{ij}, $j = 1, \ldots, d$, $i = 1, \ldots, n$..

1. Calculate an empirical dependence measure $\hat{\delta}_{jk}$ for all possible variable pairs $\{jk\}$ ($\rightarrow$ edge weights).
2. Select the tree on all nodes that maximizes the sum of absolute empirical dependencies ($\rightarrow$ maximum spanning tree):
$$\max \sum_{\substack{\text{edges } e=\{j,k\} \text{ in} \\ \text{spanning tree}}} |\hat{\delta}_{jk}|.$$
Parsimonious model selection: choose independence copula if possible.
3. For each edge $\{j, k\}$ in the selected spanning tree, select a copula and estimate the corresponding parameter(s).

Treewise construction of R-vines (Tree 1)

Dißmann et al. (2013): Capture strong dependencies of data x_{ij}, $j = 1, ..., d,\ i = 1, ..., n.$.

1. Calculate an empirical dependence measure $\hat{\delta}_{jk}$ for all possible variable pairs $\{jk\}$ ($\rightarrow$ edge weights).
2. Select the tree on all nodes that maximizes the sum of absolute empirical dependencies ($\rightarrow$ maximum spanning tree):

$$\max \sum_{\substack{\text{edges } e=\{j,k\} \text{ in} \\ \text{spanning tree}}} |\hat{\delta}_{jk}|.$$

   Parsimonious model selection: choose independence copula if possible.

3. For each edge $\{j, k\}$ in the selected spanning tree, select a copula and estimate the corresponding parameter(s).
4. Then transform to pseudo observations $F_{j|k}(x_{ij}|x_{ik}; \hat{\theta}_{jk})$ and $F_{k|j}(x_{ik}|x_{ij}; \hat{\theta}_{jk})$, $i = 1, ..., n$.

Treewise construction of R-vines (Trees $2, \ldots, d-1$)

For $\ell = 2, \ldots, d-1$:

1. Calculate the empirical dependence measure $\hat{\delta}_{jk|D}$ for all conditional variable pairs $\{j, k|D\}$ that can be part of tree T_ℓ, i.e., all edges fulfilling the proximity condition.

Treewise construction of R-vines (Trees $2, \ldots, d-1$)

For $\ell = 2, \ldots, d-1$:

1. Calculate the empirical dependence measure $\hat{\delta}_{jk|D}$ for all conditional variable pairs $\{j, k|D\}$ that can be part of tree T_ℓ, i.e., all edges fulfilling the proximity condition.
2. Among these edges, select the spanning tree that maximizes the sum of absolute empirical dependencies, i.e.,

$$\max \sum_{\substack{\text{edges } e=\{j,k|D\} \text{ in} \\ \text{spanning tree}}} |\hat{\delta}_{j,k|D}|.$$

Treewise construction of R-vines (Trees $2, \ldots, d-1$)

For $\ell = 2, \ldots, d-1$:

1. Calculate the empirical dependence measure $\hat{\delta}_{jk|D}$ for all conditional variable pairs $\{j, k|D\}$ that can be part of tree T_ℓ, i.e., all edges fulfilling the proximity condition.
2. Among these edges, select the spanning tree that maximizes the sum of absolute empirical dependencies, i.e.,
$$\max \sum_{\substack{\text{edges } e=\{j,k|D\} \text{ in} \\ \text{spanning tree}}} |\hat{\delta}_{j,k|D}|.$$
3. For each edge $\{j, k|D\}$ in the selected spanning tree, select a conditional copula and estimate the corresponding parameter(s).

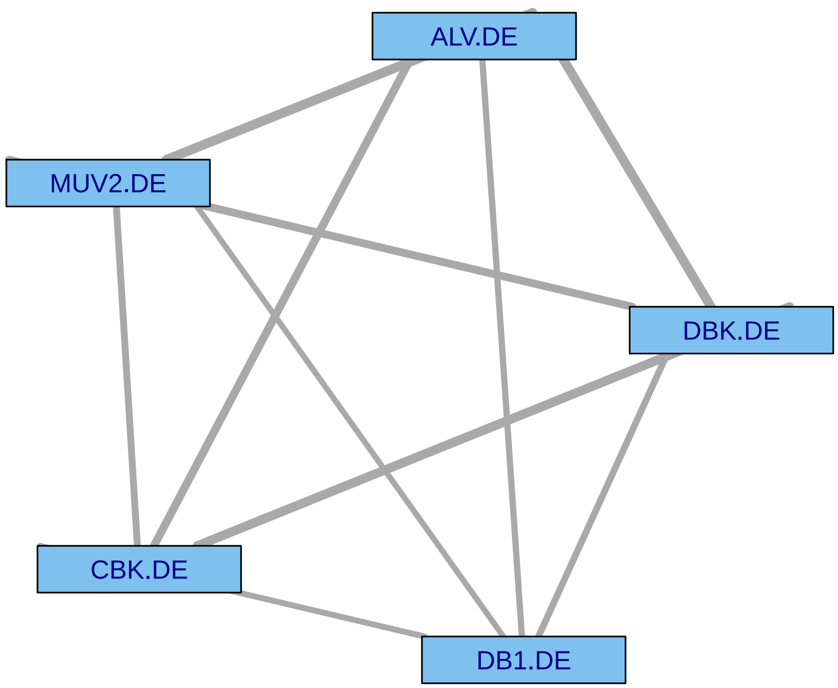
Treewise construction of R-vines (Trees $2, \ldots, d-1$)

For $\ell = 2, \ldots, d-1$:

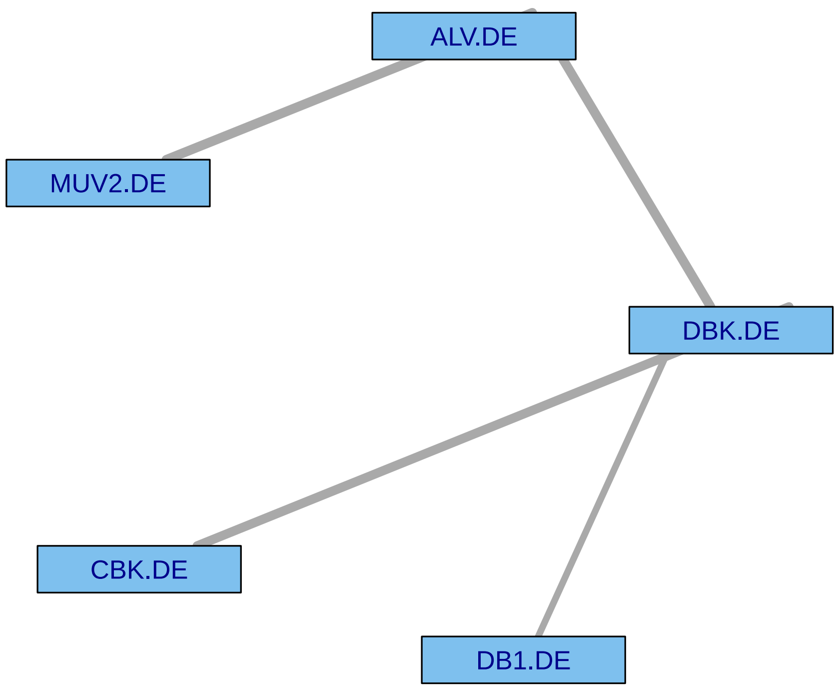
1. Calculate the empirical dependence measure $\hat{\delta}_{jk|D}$ for all conditional variable pairs $\{j, k|D\}$ that can be part of tree T_ℓ, i.e., all edges fulfilling the proximity condition.
2. Among these edges, select the spanning tree that maximizes the sum of absolute empirical dependencies, i.e.,
$$\max \sum_{\substack{\text{edges } e=\{j,k|D\} \text{ in} \\ \text{spanning tree}}} |\hat{\delta}_{j,k|D}|.$$
3. For each edge $\{j, k|D\}$ in the selected spanning tree, select a conditional copula and estimate the corresponding parameter(s).
4. Then transform to pseudo observations $F_{j|k\cup D}(x_{ij}|\boldsymbol{x}_{i,k\cup D}; \hat{\theta}_{j,k|D})$ and $F_{k|j\cup D}(x_{ik}|\boldsymbol{x}_{i,j\cup D}; \hat{\theta}_{j,k|D})$, $i = 1, \ldots, n$.

Maximum dependence tree

(1) Pairwise dependencies.



(2) Maximum dependence tree.



Vine tree selection

```
> RVineStructureSelect(data, familyset, type,
+                      selectioncrit, indeptest,
+                      level, trunclevel)
```

Vine tree selection

```
> RVineStructureSelect(data, familyset, type="RVine",
+                      selectioncrit, indeptest,
+                      level, trunclevel)
```

- R- and C-vine copulas can be selected.

Vine tree selection

```
> RVineStructureSelect(data, familyset, type,
+                      selectioncrit, indeptest,
+                      level, trunclevel=2)
```

- R- and C-vine copulas can be selected.
- The vine copula can be truncated to reduce the model complexity.

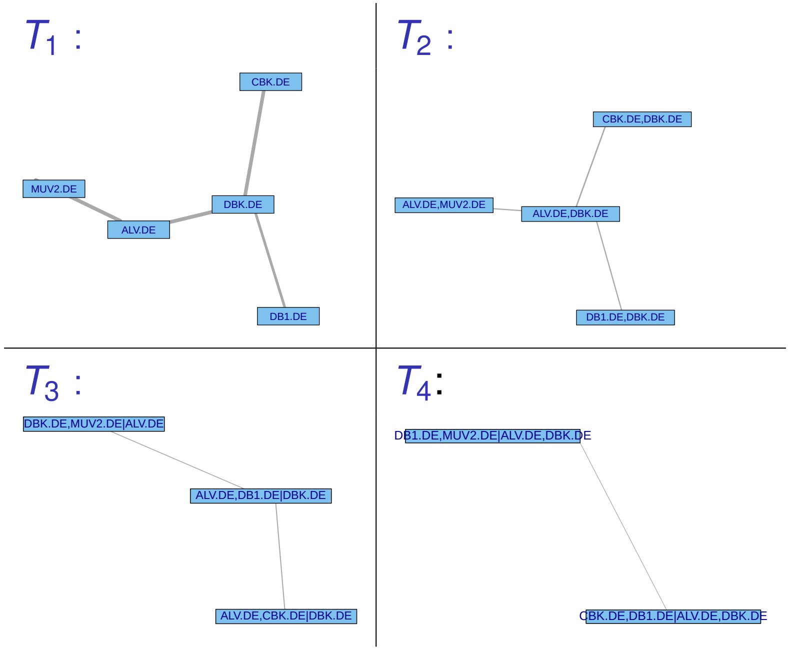
Vine tree selection

```
> RVineStructureSelect(data, familyset, type,
+                      selectioncrit, indeptest,
+                      level, trunclevel=2)
```

- R- and C-vine copulas can be selected.
- The vine copula can be truncated to reduce the model complexity.
- Illustrating R-vine copula models

```
> RVineTreePlot(data=NULL, RVM=rvm, tree=1,
+               edge.labels=c("family","theotau"))
```

Data example [VineCopula: RVineTreePlot]



Remarks

- For D-vines the path on all nodes with maximum sum of pairwise dependencies, a maximal Hamiltonian path, has be to found, i.e. a traveling salesman problem has to be solved.
- For C-vines nodes with maximum sum of pairwise dependencies to all other nodes are selected as root nodes (Czado et al. 2012).
- Kurowicka (2011) builds trees starting from the last tree to the top tree (bottom-up approach) by using empirical partial correlations as approximate measure of pairwise dependence.
- A first comparison of sequential R-vine selection methods are in Czado, Jeske, and Hofmann (2012)

Comparing vine copulas

Given competing vine copulas $\mathcal{C} = \{C_1, ..., C_m\}$ for data $\{\boldsymbol{x}_i = (x_{i1}, \ldots, x_{id}), i = 1, ..., n\}$. Which is the "best" model?

In other words: Which model C^* is statistically superior to the others?

Comparing vine copulas

Given competing vine copulas $\mathcal{C} = \{C_1, ..., C_m\}$ for data $\{\boldsymbol{x}_i = (x_{i1}, \ldots, x_{id}), i = 1, ..., n\}$. Which is the "best" model?

In other words: Which model C^* is statistically superior to the others?

Comparison using information criteria

$$C^*_{\text{AIC}} = \operatorname*{argmin}_{j \leq m} \text{AIC}(C_j) \qquad \text{or} \qquad C^*_{\text{BIC}} = \operatorname*{argmin}_{j \leq m} \text{BIC}(C_j),$$

where
$\text{AIC}(C) := -2 \sum_{i=1}^{n} \log(c(F_1(x_{i1}), \ldots, F_d(x_{id}); \theta)) + 2k_\theta$, and
$\text{BIC}(C) := -2 \sum_{i=1}^{n} \log(c(F_1(x_{i1}), \ldots, F_d(x_{id}); \theta)) + \log(n)k_\theta$,
where k_θ is the number of model parameters.

- No statement whether significantly superior to the other models!
- Problematic, when models are non-nested!

Non-nested model comparison: Vuong test

Vuong (1989) test

Compare two competing non-nested models f_1 and f_2 by their pointwise likelihoods: for i.i.d. $\boldsymbol{X}_i,\ i = 1, ..., n$, define $M_i := \log \left[\frac{f_1(\boldsymbol{X}_i|\hat{\theta}_1)}{f_2(\boldsymbol{X}_i|\hat{\theta}_2)} \right]$.

$$H_0 : \ E(M_i) = 0 \ \forall i = 1, .., n$$

For observed $M_i = m_i$, reject H_0 and prefer model 1 to model 2 at level α if

$$v := \frac{\frac{1}{n}\sum_{i=1}^{n} m_i}{\sqrt{\sum_{i=1}^{n}(m_i - \overline{m})^2}} > \Phi^{-1}\left(1 - \frac{\alpha}{2}\right).$$

Choose model 2 if $v < -\Phi^{-1}\left(1 - \frac{\alpha}{2}\right)$. No decision if $|v| \leq \Phi^{-1}\left(1 - \frac{\alpha}{2}\right)$.

Vuong test with Akaike/Schwarz correction

The Vuong test does not take into account the possibly different number of parameters of both models. Hence the test is called unadjusted and Vuong (1989) gives the definition of an adjusted statistic.

Adjusted test statistics

Let k_1 and k_2 denote the number of parameters of Models 1 and 2, respectively. Then, the Akaike and Schwarz correction for the Vuong test statistic are given by

$$\nu_{\text{Akaike}} := \frac{\frac{1}{n}\left(\sum_{i=1}^{n} m_i - (k_1 - k_2)\right)}{\sqrt{\sum_{i=1}^{n}(m_i - \overline{m})^2}},$$

and

$$\nu_{\text{Schwarz}} := \frac{\frac{1}{n}\left(\sum_{i=1}^{n} m_i - \frac{1}{2}\log(n)(k_1 - k_2)\right)}{\sqrt{\sum_{i=1}^{n}(m_i - \overline{m})^2}}.$$

Data example [VineCopula: RVineAIC/BIC, RVineVuongTest]

	log lik.	#par.	AIC	BIC
Student's t copula	1561.49	11	-3100.98	-3045.38
Student's t R-vine (seq. est.)	1589.91	20	-3139.81	-3038.72
Student's t R-vine (MLE)	1590.49	20	-3140.98	-3039.89

Data example [VineCopula: RVineAIC/BIC, RVineVuongTest]

	log lik.	#par.	AIC	BIC
Student's t copula	1561.49	11	-3100.98	-3045.38
Student's t R-vine (seq. est.)	1589.91	20	-3139.81	-3038.72
Student's t R-vine (MLE)	1590.49	20	-3140.98	-3039.89

Student's t R-vine has too many parameters! $\rightarrow$ Try Gaussian R-vine.

- Vuong test: Student's t R-vine vs. Gaussian R-vine:

$$v = 5.61 \qquad v_{\text{Akaike}} = 5.16 \qquad v_{\text{Schwarz}} = 4.00$$

$$\text{p-value} < 0.01 \quad \text{p-value}_{\text{Akaike}} < 0.01 \quad \text{p-value}_{\text{Schwarz}} < 0.01$$

$\Rightarrow$ Student's t R-vine $>$ Gaussian R-vine

Data example [VineCopula: RVineAIC/BIC, RVineVuongTest]

	log lik.	#par.	AIC	BIC
Student's t copula	1561.49	11	-3100.98	-3045.38
Student's t R-vine (seq. est.)	1589.91	20	-3139.81	-3038.72
Student's t R-vine (MLE)	1590.49	20	-3140.98	-3039.89

Student's t R-vine has too many parameters! $\rightarrow$ Try Gaussian R-vine.

- Vuong test: Student's t R-vine vs. Gaussian R-vine:

$$\nu = 5.61 \qquad \nu_{\text{Akaike}} = 5.16 \qquad \nu_{\text{Schwarz}} = 4.00$$

$$\text{p-value} < 0.01 \quad \text{p-value}_{\text{Akaike}} < 0.01 \quad \text{p-value}_{\text{Schwarz}} < 0.01$$

$\Rightarrow$ Student's t R-vine $>$ Gaussian R-vine

Need for R-vine distributions with mixed pair-copulas!

Current research

- D-vine based quantile regression (Kraus and Czado 2016)
- Examination of the simplifying assumption (Killiches, Kraus, and Czado 2016)
- Geo-spatial dependent R-vines (Erhardt, Czado, and Schepsmeier 2015)
- nonparametric vines (Nagler and Czado 2015)
- sparse vines (Müller and Czado 2016)

and many more...

Final remarks

- Vines provide a computationally tractable and highly flexible class of distributions.
- Useful for many applications in risk management such as stress testing or Value-at-Risk estimation.
- Careful marginal modeling necessary.

Final remarks

- Vines provide a computationally tractable and highly flexible class of distributions.
- Useful for many applications in risk management such as stress testing or Value-at-Risk estimation.
- Careful marginal modeling necessary.

R-package on CRAN:

- `VineCopula`: Statistical inference of R-vine copulas
  `http://cran.r-project.org/web/packages/VineCopula/`

Final remarks

- Vines provide a computationally tractable and highly flexible class of distributions.
- Useful for many applications in risk management such as stress testing or Value-at-Risk estimation.
- Careful marginal modeling necessary.

R-package on CRAN:

- `VineCopula`: Statistical inference of R-vine copulas
  `http://cran.r-project.org/web/packages/VineCopula/`

Vine resource page: `http://www.vine-copula.org`

I like to thank Claudia Czado, Daniel Kraus, Matthias Scherer as well as all my other collaborators and colleagues!

Bibliography I

Acar, E. F., C. Genest, and J. Nešlehová (2012).
Beyond simplified pair-copula constructions.
Journal of Multivariate Analysis 110, 74–90.

Czado, C., S. Jeske, and M. Hofmann (2012).
Selection strategies for regular vine copulae.
Journal de la Société Francaise de Statistique 154(1), 174–191.

Czado, C., U. Schepsmeier, and A. Min (2012).
Maximum likelihood estimation of mixed C-vines with application to exchange rates.
Statistical Modelling 12(3), 229–255.

Devroye, L. (1986).
Non-Uniform Random Variate Generation.
New York: Springer.

Dißmann, J., E. Brechmann, C. Czado, and D. Kurowicka (2013).
Selecting and estimating regular vine copulae and application to financial returns.
Computational Statistics and Data Analysis 59(1), 52 – 69.

Bibliography II

Erhardt, T. M., C. Czado, and U. Schepsmeier (2015).
Spatial composite likelihood inference using local C-vines.
Journal of Multivariate Analysis 138, 74–88.

Genest, C. and A. Favre (2007).
Everything you always wanted to know about copula modeling but were afraid to ask.
Journal of Hydrologic Engineering 12, 347–368.

Genest, C., K. Ghoudi, and L.-P. Rivest (1995).
A semiparametric estimation procedure of dependence parameters in multivariate families of distributions.
Biometrika 82, 543–552.

Hobæk Haff, I. (2013).
Parameter estimation for pair-copula constructions.
Bernoulli 19(2), 462–491.

Hobæk Haff, I., K. Aas, and A. Frigessi (2010).
On the simplified pair-copula construction – simply useful or too simplistic?
Journal of Multivariate Analysis 101(5), 1296–1310.

Bibliography III

Joe, H. and J. Xu (1996).
The estimation method of inference functions for margins for multivariate models.
Technical Report 166, Department of Statistics, University of British Columbia.

Killiches, M., D. Kraus, and C. Czado (2016).
Examination and visualization of the simplifying assumption for vine copulas in three dimensions.
arXiv preprint arXiv:1602.05795.

Kraus, D. and C. Czado (2016).
D-vine copula based quantile regression.
arXiv preprint arXiv:1510.04161.

Kurowicka, D. (2011).
Optimal truncation of vines.
In D. Kurowicka and H. Joe (Eds.), Dependence Modeling - Handbook on Vine Copulae. Singapore: World Scientific Publishing Co.

Morales-Nápoles, O. (2011).
Counting vines.
In D. Kurowicka and H. Joe (Eds.), DEPENDENCE MODELING: Vine Copula Handbook. World Scientific Publishing Co.

Bibliography IV

Morales-Nápoles, O., R. Cooke, and D. Kurowicka (2010).
About the number of vines and regular vines on n nodes.
Preprint.

Müller, D. and C. Czado (2016).
Representing sparse Gaussian DAGs as sparse R-vines allowing for non-Gaussian dependence.
arXiv preprint arXiv:1604.04202.

Nagler, T. and C. Czado (2015).
Evading the curse of dimensionality in multivariate kernel density estimation with simplified vines.
arXiv preprint arXiv:1503.03305.

Stöber, J., C. Czado, and H. Joe (2013).
Simplified pair copula constructions - limitations and extensions.
Journal of Multivariate Analysis. 119, 101–118.

Stöber, J. and U. Schepsmeier (2013).
Estimating standard errors in regular vine copula models.
Computational Statistics online first: http://link.springer.com/article/10.1007/s00180-013-0423-8.

Bibliography V

Takahashi, K. (1965).
Note on the multivariate Burr's distribution.
Annals of the Institute of Statistical Mathematics 17, 257–260.

Vuong, Q. H. (1989).
Likelihood ratio tests for model selection and non-nested hypotheses.
Econometrica 57, 307–333.

Aleksey Min

(based on the course materials of
Claudia Czado, Daniel Kraus and Matthias Scherer)

Technical University of Munich
Faculty of Mathematics

Winter term 2018/19