

7 Copulas and dependence

7.1 Copulas

7.2 Dependence concepts and measures

7.3 Normal mixture copulas

7.4 Archimedean copulas

7.5 Fitting copulas to data

7.6 A copulas-based proof of subadditivity of ES

7.1 Copulas

- We now look more closely at modelling the dependence among the components of a random vector $\mathbf{X} \sim F$ (risk-factor changes).
- **In short:** F “=” marginal dfs F_1, \dots, F_d “+” dependence structure C
- **Advantages:**
 - ▶ Most natural in a static distributional context (no time dependence; apply, e.g. to residuals of an ARMA-GARCH model)
 - ▶ Copulas allow us to understand and study dependence independently of the margins (first part of Sklar’s Theorem; see later)
 - ▶ Copulas allow for a bottom-up approach to multivariate model building (second part of Sklar’s Theorem; see later). This is often useful for constructing tailored F , e.g. when we have more information about the margins than C or for stress testing purposes.

7.1.1 Basic properties

Definition 7.1 (Copula)

A *copula* C is a df with $U(0, 1)$ margins.

Characterization

$C : [0, 1]^d \rightarrow [0, 1]$ is a copula if and only if

1) C is *grounded*, that is,

$$C(u_1, \dots, u_d) = 0 \text{ if } u_j = 0 \text{ for at least one } j \in \{1, \dots, d\}.$$

2) C has standard *uniform* univariate *margins*, that is,

$$C(1, \dots, 1, u_j, 1, \dots, 1) = u_j \text{ for all } u_j \in [0, 1] \text{ and } j \in \{1, \dots, d\}.$$

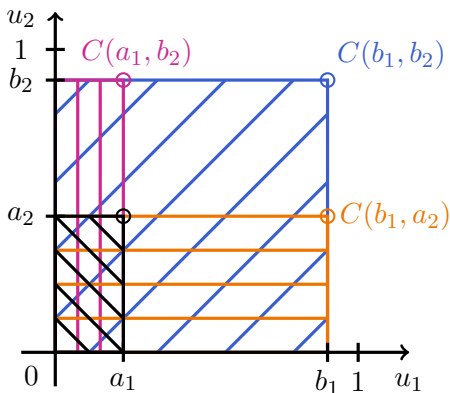
3) C is *d-increasing*, that is, for all $\mathbf{a}, \mathbf{b} \in [0, 1]^d$, $\mathbf{a} \leq \mathbf{b}$,

$$\Delta_{(\mathbf{a}, \mathbf{b}]} C = \sum_{\mathbf{i} \in \{0, 1\}^d} (-1)^{\sum_{j=1}^d i_j} C(a_1^{i_1} b_1^{1-i_1}, \dots, a_d^{i_d} b_d^{1-i_d}) \geq 0.$$

Equivalently (if existent): density $c(\mathbf{u}) \geq 0$ for all $\mathbf{u} \in (0, 1)^d$.

2-increasingness explained in a picture:

$$\begin{aligned}\Delta_{(a,b]}C &= C(b_1, b_2) - C(b_1, a_2) - C(a_1, b_2) + C(a_1, a_2) \\ &= \mathbb{P}(U \in (a, b]) \stackrel{!}{\geq} 0\end{aligned}$$



$\Rightarrow \Delta_{(a,b]}C$ is the **probability of** a random vector $U \sim C$ to be in $(a, b]$.

Preliminaries

Lemma 7.2 (Probability transformation)

Let $X \sim F$, F continuous. Then $F(X) \sim U(0, 1)$.

Idea of the proof. $\mathbb{P}(F(X) \leq u) = \mathbb{P}(F^{\leftarrow}(F(X)) \leq F^{\leftarrow}(u)) = \mathbb{P}(X \leq F^{\leftarrow}(u)) = F(F^{\leftarrow}(u)) = u$, $u \in [0, 1]$; more details in the appendix. \square

Note that F needs to be **continuous** (otherwise $F(X)$ would not reach all intervals $\subseteq [0, 1]$).

Lemma 7.3 (Quantile transformation)

Let $U \sim U(0, 1)$ and F be any df. Then $X = F^{\leftarrow}(U) \sim F$.

Proof. $\mathbb{P}(F^{\leftarrow}(U) \leq x) \stackrel{(G15)}{=} \mathbb{P}(U \leq F(x)) = F(x)$, $x \in \mathbb{R}$. \square

Probability and quantile transformations are the key to all applications involving copulas. They allow us to go from \mathbb{R}^d to $[0, 1]^d$ and back.

Sklar's Theorem

Theorem 7.4 (Sklar's Theorem)

- 1) For any df F with margins F_1, \dots, F_d , there exists a copula C such that

$$F(x_1, \dots, x_d) = C(F_1(x_1), \dots, F_d(x_d)), \quad \mathbf{x} \in \mathbb{R}^d. \quad (26)$$

C is uniquely defined on $\prod_{j=1}^d \text{ran } F_j$ and given by

$$C(u_1, \dots, u_d) = F(F_1^{\leftarrow}(u_1), \dots, F_d^{\leftarrow}(u_d)), \quad \mathbf{u} \in \prod_{j=1}^d \text{ran } F_j,$$

where $\text{ran } F_j = \{F_j(x) : x \in \mathbb{R}\}$ denotes the *range* of F_j .

- 2) Conversely, given any copula C and univariate dfs F_1, \dots, F_d , F defined by (26) is a df with margins F_1, \dots, F_d .

Proof.

- 1) **Proof for continuous F_1, \dots, F_d only.** Let $\mathbf{X} \sim F$ and define $U_j = F_j(X_j)$, $j \in \{1, \dots, d\}$. By the probability transformation, $U_j \sim U(0, 1)$ (continuity!), $j \in \{1, \dots, d\}$, so the df C of \mathbf{U} is a copula. Since $F_j \uparrow$ on $\text{ran } X_j$, (G13) implies that $X_j = F_j^{\leftarrow}(F_j(X_j)) = F_j^{\leftarrow}(U_j)$, $j \in \{1, \dots, d\}$. Therefore,

$$\begin{aligned} F(\mathbf{x}) &= \mathbb{P}(X_j \leq x_j \ \forall j) = \mathbb{P}(F_j^{\leftarrow}(U_j) \leq x_j \ \forall j) \stackrel{\text{(G15)}}{=} \mathbb{P}(U_j \leq F_j(x_j) \ \forall j) \\ &= C(F_1(x_1), \dots, F_d(x_d)), \quad \mathbf{x} \in \mathbb{R}^d. \end{aligned}$$

Hence C is a copula and satisfies (26).

(G14) implies that $F_j(F_j^{\leftarrow}(u_j)) = u_j$ for all $u_j \in \text{ran } F_j$, so

$$\begin{aligned} C(u_1, \dots, u_d) &= C(F_1(F_1^{\leftarrow}(u_1)), \dots, F_d(F_d^{\leftarrow}(u_d))) \\ &\stackrel{(26)}{=} F(F_1^{\leftarrow}(u_1), \dots, F_d^{\leftarrow}(u_d)), \quad \mathbf{u} \in \prod_{j=1}^d \text{ran } F_j. \end{aligned}$$

2) For $U \sim C$, define $\mathbf{X} = (F_1^{\leftarrow}(U_1), \dots, F_d^{\leftarrow}(U_d))$. Then

$$\begin{aligned}\mathbb{P}(\mathbf{X} \leq \mathbf{x}) &= \mathbb{P}(F_j^{\leftarrow}(U_j) \leq x_j \ \forall j) \stackrel{\text{(GI5)}}{=} \mathbb{P}(U_j \leq F_j(x_j) \ \forall j) \\ &= C(F_1(x_1), \dots, F_d(x_d)), \quad \mathbf{x} \in \mathbb{R}^d.\end{aligned}$$

Therefore, F defined by (26) is a df (that of \mathbf{X}), with margins F_1, \dots, F_d (obtained by the quantile transformation). \square

Example 7.5 (Bivariate Bernoulli distribution)

Let (X_1, X_2) follow a bivariate Bernoulli distribution with $\mathbb{P}(X_1 = k, X_2 = l) = 1/4$, $k, l \in \{0, 1\}$. $\Rightarrow \mathbb{P}(X_j = k) = 1/2$, $k \in \{0, 1\}$, $\text{ran } F_j = \{0, 1/2, 1\}$, $j \in \{1, 2\}$. Any copula with $C(1/2, 1/2) = 1/4$ satisfies (26) (e.g. $C(u_1, u_2) = \Pi(u_1, u_2)$ or the diagonal copula $C(u_1, u_2) = \min\{u_1, u_2, (\delta(u_1) + \delta(u_2))/2\}$ with $\delta(u) = u^2$).

- A copula model for \mathbf{X} means $F(\mathbf{x}) = C(F_1(x_1), \dots, F_d(x_d))$ for some (parametric) copula C and (parametric) marginals F_1, \dots, F_d .
- \mathbf{X} (or F) with margins F_1, \dots, F_d has copula C if (26) holds.

Invariance principle

Lemma 7.6 (Core of the invariance principle)

Let $X_j \sim F_j$, F_j continuous, $j \in \{1, \dots, d\}$. Then

$$\mathbf{X} \text{ has copula } C \iff (F_1(X_1), \dots, F_d(X_d)) \sim C.$$

Proof. See the appendix. □

Theorem 7.7 (Invariance principle)

Let $\mathbf{X} \sim F$ with continuous margins F_1, \dots, F_d and copula C . If $T_j \uparrow$ on $\text{ran } X_j$ for all j , then $(T_1(X_1), \dots, T_d(X_d))$ (also) has copula C .

Proof. W.l.o.g. assume T_j to be right-continuous at its at most countably many discontinuities (since X_j is continuously distributed, we only change $T_j(X_j)$ on a null set). Since $T_j \uparrow$ on $\text{ran } X_j$ and X_j is continuously distributed, $T_j(X_j)$ is continuously distributed and we have

$$\begin{aligned}
 F_{T_j(X_j)}(x) &= \mathbb{P}(T_j(X_j) \leq x) = \mathbb{P}(T_j(X_j) < x) \stackrel{(G15)}{=} \mathbb{P}(X_j < T_j^{\leftarrow}(x)) \\
 &= \mathbb{P}(X_j \leq T_j^{\leftarrow}(x)) = F_j(T_j^{\leftarrow}(x)), \quad x \in \mathbb{R}.
 \end{aligned}$$

This implies that $\mathbb{P}(F_{T_j(X_j)}(T_j(X_j)) \leq u_j \forall j)$ equals

$$\mathbb{P}(F_j(T_j^{\leftarrow}(T_j(X_j))) \leq u_j \forall j) \stackrel{(G13)}{=} \mathbb{P}(F_j(X_j) \leq u_j \forall j) \stackrel{\text{L.7.6}}{=} \text{"only if"} C(\mathbf{u}).$$

The claim follows from the if part (" \Leftarrow ") of Lemma 7.6. □

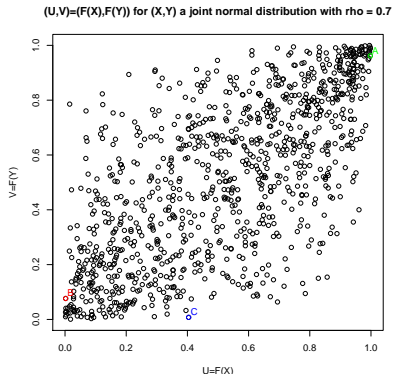
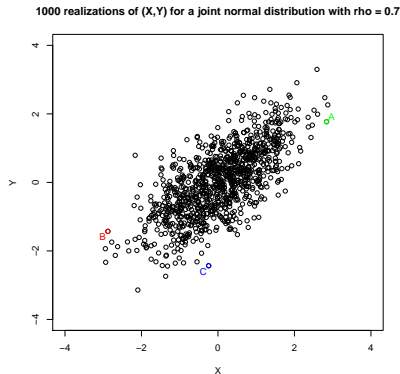
Interpretation of Sklar's Theorem (and the invariance principle)

- 1) Part 1) of Sklar's Theorem allows one to **decompose any df F into its margins and a copula**. This, together with the invariance principle, allows one to **study dependence independently of the margins via the margin-free $\mathbf{U} = (F_1(X_1), \dots, F_d(X_d))$ instead of $\mathbf{X} = (X_1, \dots, X_d)$** (they both have the same copula!). This is interesting for statistical applications, e.g. **parameter estimation** or **goodness-of-fit**.
- 2) Part 2) allows one to **construct flexible multivariate distributions** for particular applications.

Visualizing the first part of Sklar's Theorem

Left: Scatter plot of $n = 1000$ samples from $(X_1, X_2) \sim N_2(\mathbf{0}, P)$, where $P = \begin{pmatrix} 1 & 0.7 \\ 0.7 & 1 \end{pmatrix}$. We mark three points A, B, C.

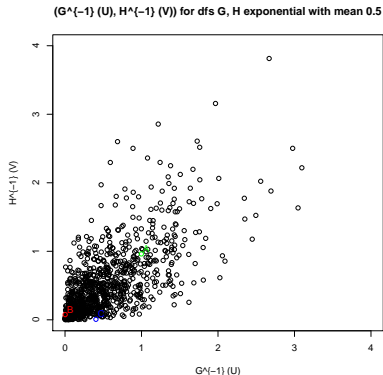
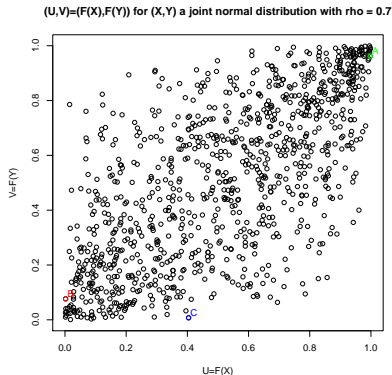
Right: Scatter plot of the corresponding Gauss copula (after applying the df Φ of $N(0, 1)$). Note how A, B, C change.



Visualizing the second part of Sklar's Theorem

Left: Same Gauss copula scatter plot as before. Apply marginal Exp(2)-quantile functions ($F_j^{-1}(u) = -\log(1-u)/2$, $j \in \{1, 2\}$).

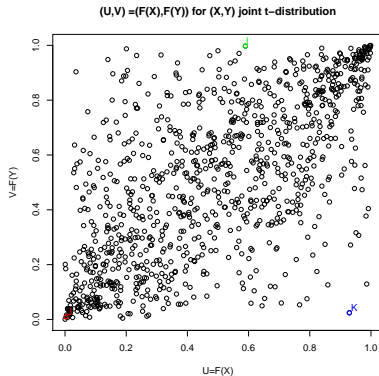
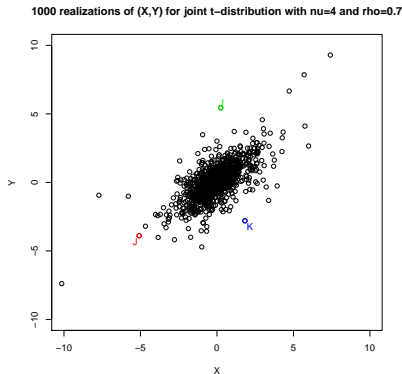
Right: The corresponding transformed random variates. Again, note the three points A, B, C.



Visualizing the first part of Sklar's Theorem

Left: Scatter plot of $n = 1000$ samples from $(X_1, X_2) \sim t_2(4, \mathbf{0}, P)$, where $P = \begin{pmatrix} 1 & 0.7 \\ 0.7 & 1 \end{pmatrix}$. We mark three points **I**, **J**, **K**.

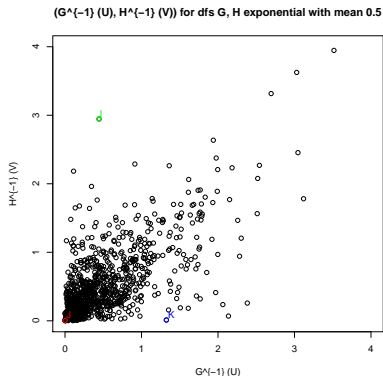
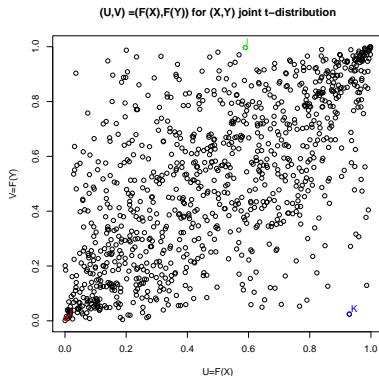
Right: Scatter plot of the corresponding t_4 copula (after applying the df t_4). Note how **A**, **B**, **C** change.



Visualizing the second part of Sklar's Theorem

Left: Same t_4 copula scatter plot as before. Apply marginal Exp(2)-quantile functions ($F_j^{-1}(u) = -\log(1-u)/2$, $j \in \{1, 2\}$).

Right: The corresponding transformed random variates. Again, note the three points I, J, K.



Fréchet–Hoeffding bounds

Theorem 7.8 (Fréchet–Hoeffding bounds)

Let $W(\mathbf{u}) = \max\{\sum_{j=1}^d u_j - d + 1, 0\}$ and $M(\mathbf{u}) = \min_{1 \leq j \leq d} \{u_j\}$.

1) For any d -dimensional copula C ,

$$W(\mathbf{u}) \leq C(\mathbf{u}) \leq M(\mathbf{u}), \quad \mathbf{u} \in [0, 1]^d.$$

2) W is a copula if and only if $d = 2$.

3) M is a copula for all $d \geq 2$.

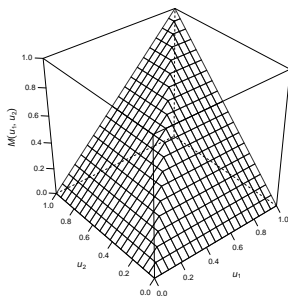
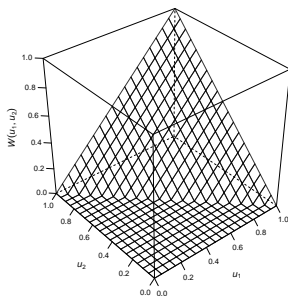
Proof. See the appendix. □

■ It is easy to verify that, for $U \sim U(0, 1)$,

▶ $(U, \dots, U) \sim M$;

▶ $(U, 1 - U) \sim W$.

- Plot of W, M for $d = 2$ (compare with $(U, 1 - U) \sim W$, $(U, U) \sim M$)



- The Fréchet–Hoeffding bounds correspond to perfect dependence (negative for W ; positive for M); see Proposition 7.14 later.
- The Fréchet–Hoeffding bounds lead to bounds for any df F , via

$$\max\left\{\sum_{j=1}^d F_j(x_j) - d + 1, 0\right\} \leq F(\mathbf{x}) \leq \min_{1 \leq j \leq d} \{F_j(x_j)\}.$$

We will use them later to derive bounds for the correlation coefficient.

7.1.2 Examples of copulas

- *Fundamental copulas*: important special copulas;
- *Implicit copulas*: extracted from known F via Sklar's Theorem;
- *Explicit copulas*: have simple closed-form expressions and follow construction principles of copulas.

Fundamental copulas

- $\Pi(\mathbf{u}) = \prod_{j=1}^d u_j$ is the *independence copula* since $C(F_1(x_1), \dots, F_d(x_d)) \stackrel{\text{Sklar}}{=} F(\mathbf{x}) \stackrel{\text{ind.}}{=} \prod_{j=1}^d F_j(x_j)$ if and only if $C(\mathbf{u}) = \Pi(\mathbf{u})$ (now replace x_j by $F_j^{\leftarrow}(u_j)$ and apply (GI4)). Therefore, X_1, \dots, X_d are independent if and only if their copula is Π .
- The Fréchet–Hoeffding bound W is the *countermonotonicity copula*. It is the df of $(U, 1 - U)$. If X_1, X_2 are perfectly negatively dependent (X_2 is a.s. a strictly decreasing function in X_1), their copula is W .

- The Fréchet–Hoeffding bound M is the *comonotonicity copula*. It is the df of (U, \dots, U) . If X_1, \dots, X_d are perfectly positively dependent (X_2, \dots, X_{d-1} are a.s. strictly increasing functions in X_1), their copula is M .

Implicit copulas

Elliptical copulas are implicit copulas arising from elliptical distributions via Sklar's Theorem. The two most prominent parametric families in this class are the *Gauss copula* and the *t copula*.

Gauss copulas

- Consider (w.l.o.g.) $\mathbf{X} \sim N_d(\mathbf{0}, P)$. The *Gauss copula* (family) is given by

$$\begin{aligned} C_P^{\text{Ga}}(\mathbf{u}) &= \mathbb{P}(\Phi(X_1) \leq u_1, \dots, \Phi(X_d) \leq u_d) \\ &= \Phi_P(\Phi^{-1}(u_1), \dots, \Phi^{-1}(u_d)) \end{aligned}$$

where Φ_P is the df of $N_d(\mathbf{0}, P)$ and Φ the df of $N(0, 1)$.

- Special cases: If $P = I_d$ then $C = \Pi$, and if $P = J_d = \mathbf{1}\mathbf{1}'$ then $C = M$.
If $d = 2$ and $\rho = P_{12} = -1$ then $C = W$.
- Sklar's Theorem \Rightarrow The density of $C(\mathbf{u}) = F(F_1^{\leftarrow}(u_1), \dots, F_d^{\leftarrow}(u_d))$ is

$$c(\mathbf{u}) = \frac{f(F_1^{\leftarrow}(u_1), \dots, F_d^{\leftarrow}(u_d))}{\prod_{j=1}^d f_j(F_j^{\leftarrow}(u_j))}, \quad \mathbf{u} \in (0, 1)^d.$$

In particular, the density of C_P^{Ga} is

$$c_P^{\text{Ga}}(\mathbf{u}) = \frac{1}{\sqrt{\det P}} \exp\left(-\frac{1}{2} \mathbf{x}'(P^{-1} - I_d)\mathbf{x}\right), \quad (27)$$

where $\mathbf{x} = (\Phi^{-1}(u_1), \dots, \Phi^{-1}(u_d))$.

t copulas

- Consider (w.l.o.g.) $\mathbf{X} \sim t_d(\nu, \mathbf{0}, P)$. The t copula (family) is given by

$$\begin{aligned} C_{\nu, P}^t(\mathbf{u}) &= \mathbb{P}(t_\nu(X_1) \leq u_1, \dots, t_\nu(X_d) \leq u_d) \\ &= t_{\nu, P}(t_\nu^{-1}(u_1), \dots, t_\nu^{-1}(u_d)) \end{aligned}$$

where $t_{\nu,P}$ is the df of $t_d(\nu, \mathbf{0}, P)$ and t_ν the df of the univariate t distribution with ν degrees of freedom.

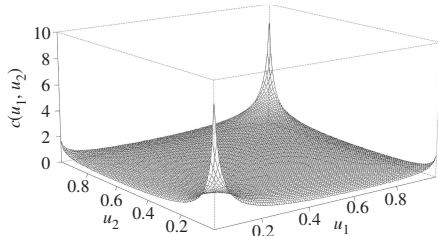
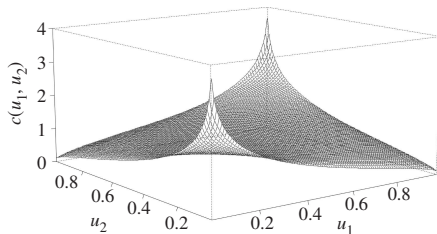
- Special cases: $P = J_d = \mathbf{1}\mathbf{1}'$ then $C = M$. However, if $P = I_d$ then $C \neq \Pi$ (unless $\nu = \infty$ in which case $C_{\nu,P}^t = C_P^{\text{Ga}}$). If $d = 2$ and $\rho = P_{12} = -1$ then $C = W$.
- Sklar's Theorem \Rightarrow The density of $C_{\nu,P}^t$ is

$$c_{\nu,P}^t(\mathbf{u}) = \frac{\Gamma((\nu + d)/2)}{\Gamma(\nu/2)\sqrt{\det P}} \left(\frac{\Gamma(\nu/2)}{\Gamma((\nu + 1)/2)} \right)^d \frac{(1 + \mathbf{x}'P^{-1}\mathbf{x}/\nu)^{-(\nu+d)/2}}{\prod_{j=1}^d (1 + x_j^2/\nu)^{-(\nu+1)/2}},$$

for $\mathbf{x} = (t_\nu^{-1}(u_1), \dots, t_\nu^{-1}(u_d))$.

- For more details, see Demarta and McNeil (2005).
- For scatter plots, see the visualization of Sklar's Theorem above. Note the difference in the tails: The smaller ν , the more mass is concentrated in the joint tails.

Perspective plots of the densities of $C'_{\rho=0.3}{}^{\text{Ga}}$ (left) and $C_{4,\rho=0.3}^t(\mathbf{u})$ (right).



Advantages and drawbacks of elliptical copulas (see later, too):

Advantages:

- Modelling pairwise dependencies (comparably flexible)
- Density available
- Sampling (typically) simple

Drawbacks:

- Typically, C is not explicit
- Radially symmetric (so the same lower/upper tail behaviour)

Explicit copulas

Archimedean copulas are copulas of the form

$$C(\mathbf{u}) = \psi(\psi^{-1}(u_1) + \cdots + \psi^{-1}(u_d)), \quad \mathbf{u} \in [0, 1]^d,$$

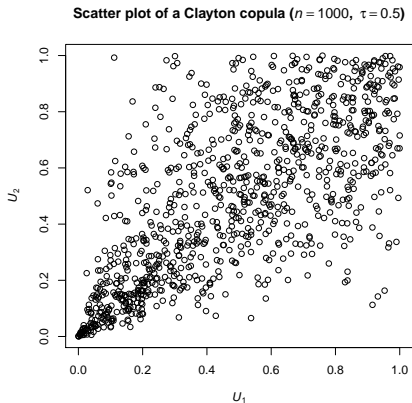
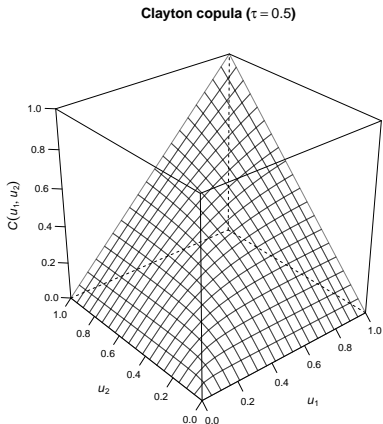
where the (*Archimedean*) generator $\psi : [0, \infty) \rightarrow [0, 1]$ is \downarrow on $[0, \inf\{t : \psi(t) = 0\}]$ and satisfies $\psi(0) = 1$, $\psi(\infty) = \lim_{t \rightarrow \infty} \psi(t) = 0$; we set $\psi^{-1}(0) = \inf\{t : \psi(t) = 0\}$. The set of all generators is denoted by Ψ . If $\psi(t) > 0$, $t \in [0, \infty)$, we call ψ *strict*.

Examples

- **Clayton copula:** Obtained for $\psi(t) = (1+t)^{-1/\theta}$, $t \in [0, \infty)$, $\theta \in (0, \infty)$
 $\Rightarrow C_{\theta}^c(\mathbf{u}) = (u_1^{-\theta} + \cdots + u_d^{-\theta} - d + 1)^{-1/\theta}$. For $\theta \downarrow 0$, $C \rightarrow \Pi$; and for $\theta \uparrow \infty$, $C \rightarrow M$.
- **Gumbel copula:** Obtained for $\psi(t) = \exp(-t^{1/\theta})$, $t \in [0, \infty)$, $\theta \in [1, \infty)$
 $\Rightarrow C_{\theta}^G(\mathbf{u}) = \exp(-((- \log u_1)^{\theta} + \cdots + (- \log u_d)^{\theta})^{1/\theta})$. For $\theta = 1$, $C = \Pi$; and for $\theta \rightarrow \infty$, $C \rightarrow M$.

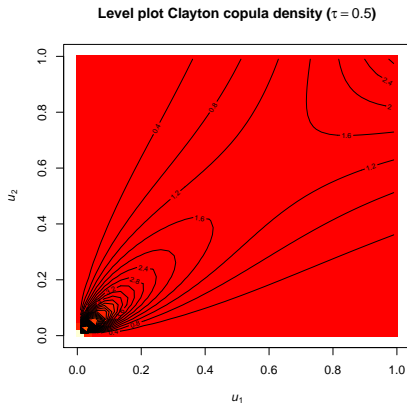
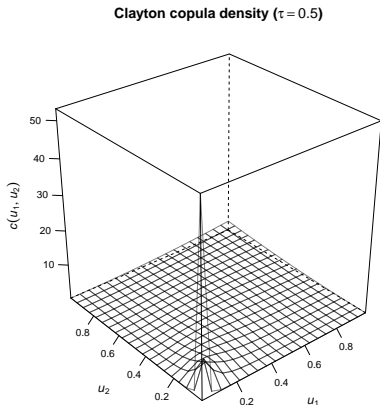
Left: Plot of a bivariate **Clayton copula** (Kendall's tau 0.5; see later).

Right: Corresponding **scatter plot** (sample size $n = 1000$)



Left: Plot of the corresponding density.

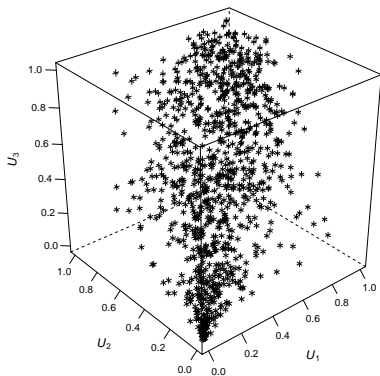
Right: Level plot of the density (with heat colors).



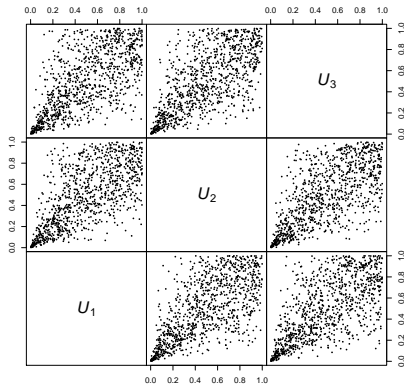
Left: Cloud plot of a trivariate Clayton copula (sample size $n = 1000$; Kendall's tau 0.5).

Right: Corresponding scatter plot matrix.

Clayton copula cloud plot ($n = 1000$, $\tau = 0.5$)

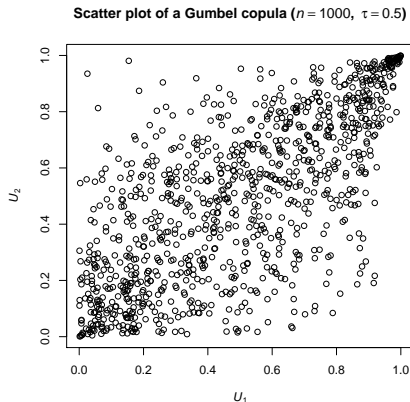
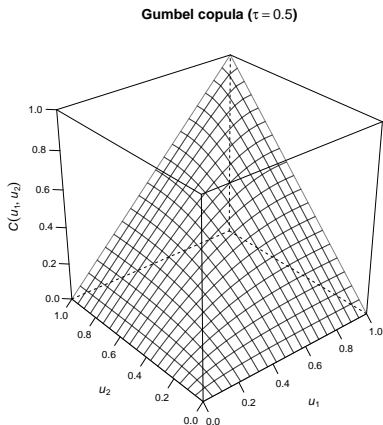


Scatter plot matrix of a Clayton copula ($n = 1000$, $\tau = 0.5$)



Left: Plot of a bivariate **Gumbel copula** (Kendall's tau 0.5).

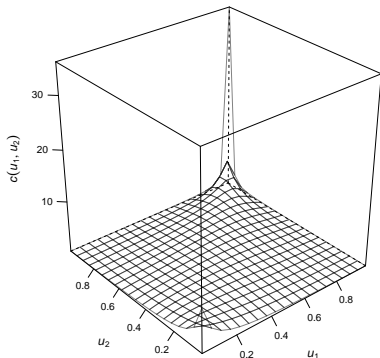
Right: Corresponding **scatter plot** (sample size $n = 1000$)



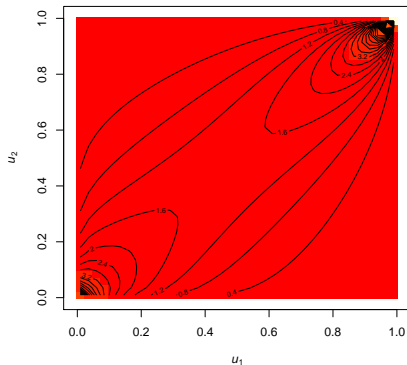
Left: Plot of the corresponding density.

Right: Level plot of the density (with heat colors).

Gumbel copula density ($\tau = 0.5$)



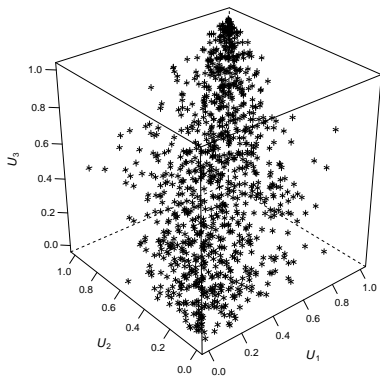
Level plot Gumbel copula density ($\tau = 0.5$)



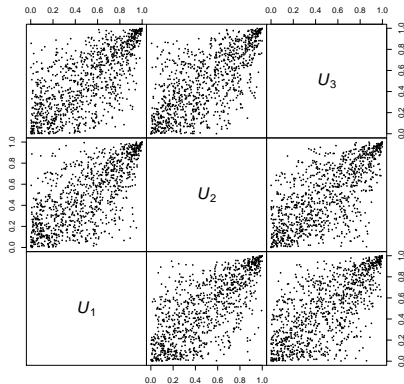
Left: Cloud plot of a trivariate Gumbel copula (sample size $n = 1000$; Kendall's tau 0.5).

Right: Corresponding scatter plot matrix.

Gumbel copula cloud plot ($n = 1000$, $\tau = 0.5$)



Scatter plot matrix of a Gumbel copula ($n = 1000$, $\tau = 0.5$)



Advantages and drawbacks of Archimedean copulas (see later, too):

Advantages:

- Typically **explicit** (if ψ^{-1} is available)
- Useful in calculations: **Properties** can typically be expressed **in terms of ψ**
- **Densities** of various examples available
- **Sampling** often simple
- **Not restricted to radial symmetry**

Drawbacks:

- All margins of the same dimension are equal (symmetry or **exchangeability**; see later)
- Often used only with a small **number of parameters** (some extensions available, but still less than $d(d-1)/2$)

7.1.3 Meta distributions

- *Fréchet class*: Class of all dfs F with given marginal dfs F_1, \dots, F_d ;
Meta- C models: All dfs F with the same given copula C .
- **Example**: A meta-Gauss model is a multivariate df F with Gauss copula C and some margins F_1, \dots, F_d .

7.1.4 Simulation of copulas and meta distributions

Sampling implicit copulas

Due to their construction via Sklar's Theorem, implicit copulas can be sampled via Lemma 7.6.

Algorithm 7.9 (Simulation of implicit copulas)

- 1) Sample $\mathbf{X} \sim F$, where F is a df with continuous margins F_1, \dots, F_d .
- 2) Return $\mathbf{U} = (F_1(X_1), \dots, F_d(X_d))$ (probability transformation).

Example 7.10

- Sampling Gauss copulas C_P^{Ga} :

- 1) Sample $\mathbf{X} \sim N_d(\mathbf{0}, P)$ ($\mathbf{X} \stackrel{d}{=} A\mathbf{Z}$ for $AA' = P$, $\mathbf{Z} \sim N_d(\mathbf{0}, I_d)$).
- 2) Return $\mathbf{U} = (\Phi(X_1), \dots, \Phi(X_d))$.

- Sampling t_ν copulas $C_{\nu, P}^t$:

- 1) Sample $\mathbf{X} \sim t_d(\nu, \mathbf{0}, P)$ ($\mathbf{X} \stackrel{d}{=} \sqrt{W}A\mathbf{Z}$ for $W = \frac{1}{V}$, $V \sim \Gamma(\frac{\nu}{2}, \frac{\nu}{2})$).
- 2) Return $\mathbf{U} = (t_\nu(X_1), \dots, t_\nu(X_d))$.

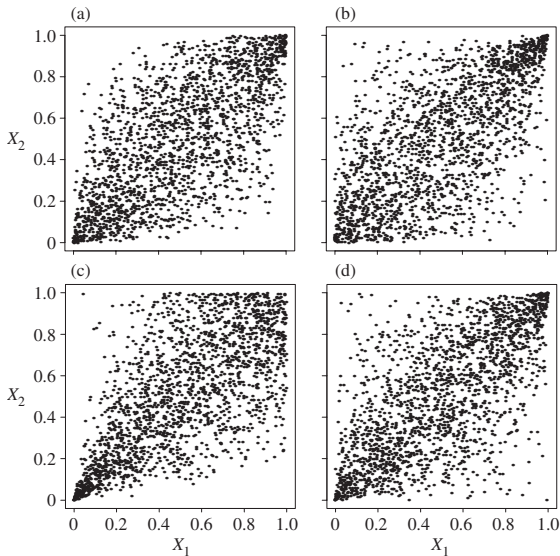
Sampling meta distributions

Meta- C distributions can be sampled via Sklar's Theorem, Part 2).

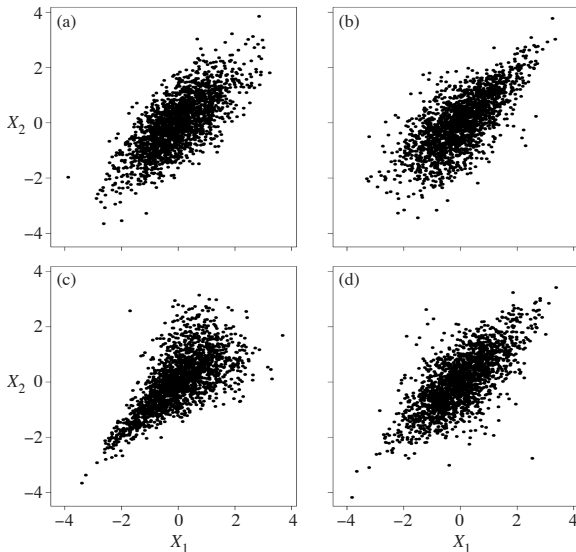
Algorithm 7.11 (Sampling)

- 1) Sample $\mathbf{U} \sim C$.
- 2) Return $\mathbf{X} = (F_1^{\leftarrow}(U_1), \dots, F_d^{\leftarrow}(U_d))$ (quantile transformation).

2000 samples from (a): $C_{\rho=0.7}^{\text{Ga}}$; (b): $C_{\theta=2}^{\text{G}}$; (c): $C_{\theta=2.2}^{\text{C}}$; (d): $C_{\nu=4, \rho=0.71}^t$



... transformed to $N(0, 1)$ margins; all have linear correlation ≈ 0.7 !



A general sampling algorithm

For a general copula C (without further information), the only known sampling algorithm is the *conditional distribution method*; see Embrechts et al. (2003) and Hofert (2010, p. 41).

Theorem 7.12 (Conditional distribution method)

If C is a d -dimensional copula and $U' \sim U(0, 1)^d$, let

$$U_1 = U'_1,$$

$$U_2 = C^{\leftarrow}(U'_2 | U_1),$$

$$\vdots$$

$$U_d = C^{\leftarrow}(U'_d | U_1, \dots, U_{d-1}).$$

Then $U \sim C$.

This typically involves numerical root-finding and the following result.

Theorem 7.13 (Schmitz (2003))

Let C be a d -dimensional copula which admits, for $d \geq 3$, continuous partial derivatives w.r.t. the first $d - 1$ arguments. Then

$$C(u_j | u_1, \dots, u_{j-1}) = \frac{D_{j-1, \dots, 1} C^{(1, \dots, j)}(u_1, \dots, u_j)}{D_{j-1, \dots, 1} C^{(1, \dots, j-1)}(u_1, \dots, u_{j-1})}$$

for a.e. $u_1, \dots, u_{j-1} \in [0, 1]$, where the superscripts denote the corresponding marginal copulas and $D_{j-1, \dots, 1}$ the differential operator w.r.t. the first $j - 1$ components.

- For $d = 2$ one obtains that $C(u_2 | u_1) = D_1 C(u_1, u_2)$ for a.e. $u_1 \in [0, 1]$.
- For most well-known copula families, the conditional distribution method is neither simple to apply nor fast \Rightarrow Efficient sampling algorithms are typically family-specific.

7.1.5 Further properties of copulas

Survival copulas

- If $U \sim C$, then $\mathbf{1} - U \sim \hat{C}$, the *survival copula* of C .
- \hat{C} can be expressed as

$$\hat{C}(\mathbf{u}) = \sum_{J \subseteq \{1, \dots, d\}} (-1)^{|J|} C((1 - u_1)^{I_J(1)}, \dots, (1 - u_d)^{I_J(d)})$$

in terms of its corresponding copula (essentially an application of the *Poincaré–Sylvester sieve formula*). For $d = 2$,

$$\begin{aligned}\hat{C}(u_1, u_2) &= 1 - (1 - u_1) - (1 - u_2) + C(1 - u_1, 1 - u_2) \\ &= -1 + u_1 + u_2 + C(1 - u_1, 1 - u_2).\end{aligned}$$

- If C admits a density, $\hat{c}(\mathbf{u}) = c(\mathbf{1} - \mathbf{u})$.
- If $\hat{C} = C$, C is called *radially symmetric*. Check that W , Π , and M are radially symmetric.

- One can show: If X_j is symmetrically distributed about a_j , $j \in \{1, \dots, d\}$, then \mathbf{X} is radially symmetric about \mathbf{a} if and only if $C = \hat{C}$.
- Sklar's Theorem can also be formulated for survival functions. In this case, the main part reads

$$\bar{F}(\mathbf{x}) = \hat{C}(\bar{F}_1(x_1), \dots, \bar{F}_d(x_d)),$$

where $F(\mathbf{x}) = \mathbb{P}(\mathbf{X} > \mathbf{x})$ with corresponding marginal survival functions $\bar{F}_1, \dots, \bar{F}_d$ (with $\bar{F}_j(x) = \mathbb{P}(X_j > x)$).

⇒ Survival copulas combine marginal survival functions to joint survival functions. Note that \hat{C} is a df, whereas \bar{F} and $\bar{F}_1, \dots, \bar{F}_d$ are not!

Copula densities

- By **Sklar's Theorem**, if F_j has density f_j , $j \in \{1, \dots, d\}$, and C has density c , then the density f of F satisfies

$$f(\mathbf{x}) = c(F_1(x_1), \dots, F_d(x_d)) \prod_{j=1}^d f_j(x_j) \quad (28)$$

As seen before, we can recover c via

$$c(\mathbf{u}) = \frac{f(F_1^{-1}(u_1), \dots, F_d^{-1}(u_d))}{f_1(F_1^{-1}(u_1)) \cdots f_d(F_d^{-1}(u_d))}.$$

- It follows from (28) that the **log-density** splits into

$$\log f(\mathbf{x}) = \log c(F_1(x_1), \dots, F_d(x_d)) + \sum_{j=1}^d \log f_j(x_j).$$

which **allows for a two-stage estimation** (**marginal** and **copula parameters**); see Section 7.5.

Exchangeability

- \mathbf{X} is *exchangeable* if

$$(X_1, \dots, X_d) \stackrel{d}{=} (X_{\pi(1)}, \dots, X_{\pi(d)})$$

for any permutation $(\pi(1), \dots, \pi(d))$ of $(1, \dots, d)$.

- A copula C is *exchangeable* if it is the df of an exchangeable \mathbf{U} with $U(0, 1)$ margins. This holds if and only if $C(u_1, \dots, u_d) = C(u_{\pi(1)}, \dots, u_{\pi(d)})$ for all possible permutations of arguments, i.e. if C is *symmetric*.
- Exchangeable/symmetric copulas are useful for approximate modelling homogeneous portfolios.
- **Examples:**
 - ▶ Archimedean copulas
 - ▶ Elliptical copulas (such as Gauss/ t) for equicorrelated P (i.e. $P = \rho J_d + (1 - \rho)I_d$ for $\rho \geq -1/(d - 1)$); in particular, $d = 2$

7.2 Dependence concepts and measures

Measures of association/dependence are scalar measures which summarize the dependence in terms of a single number. There are better and worse examples of such measures, which we will study in this section.

7.2.1 Perfect dependence

X_1, X_2 are *countermonotone* if (X_1, X_2) has copula W .

X_1, \dots, X_d are *comonotone* if (X_1, \dots, X_d) has copula M .

Proposition 7.14 (Perfect dependence)

- 1) $X_2 = T(X_1)$ a.s. with decreasing $T(x) = F_2^{\leftarrow}(1 - F_1(x))$ (countermonotone) if and only if $C(u_1, u_2) = W(u_1, u_2)$, $u_1, u_2 \in [0, 1]$.
- 2) $X_j = T_j(X_1)$ a.s. with increasing $T_j(x) = F_j^{\leftarrow}(F_1(x))$, $j \in \{2, \dots, d\}$ (comonotone), if and only if $C(\mathbf{u}) = M(\mathbf{u})$, $\mathbf{u} \in [0, 1]^d$.

Proof. See the appendix. □

Proposition 7.15 (Comonotone additivity)

Let $\alpha \in (0, 1)$ and $X_j \sim F_j$, $j \in \{1, \dots, d\}$, be comontone. Then $F_{X_1+\dots+X_d}^{\leftarrow}(\alpha) = F_1^{\leftarrow}(\alpha) + \dots + F_d^{\leftarrow}(\alpha)$; technical proof, see appendix.

7.2.2 Linear correlation

For two random variables X_1 and X_2 with $\mathbb{E}(X_j^2) < \infty$, $j \in \{1, 2\}$, the (*linear* or *Pearson's*) *correlation coefficient* ρ is defined by

$$\rho(X_1, X_2) = \frac{\text{cov}(X_1, X_2)}{\sqrt{\text{var } X_1} \sqrt{\text{var } X_2}} = \frac{\mathbb{E}((X_1 - \mathbb{E}X_1)(X_2 - \mathbb{E}X_2))}{\sqrt{\mathbb{E}((X_1 - \mathbb{E}X_1)^2)} \sqrt{\mathbb{E}((X_2 - \mathbb{E}X_2)^2)}}.$$

Proposition 7.16 (Hoeffding's identity)

Let $X_j \sim F_j$, $j \in \{1, 2\}$, be two random variables with $\mathbb{E}(X_j^2) < \infty$, $j \in \{1, 2\}$, and joint distribution function F . Then

$$\text{cov}(X_1, X_2) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (F(x_1, x_2) - F_1(x_1)F_2(x_2)) dx_1 dx_2.$$

Classical properties and drawbacks of linear correlation

Let X_1 and X_2 be two random variables with $\mathbb{E}(X_j^2) < \infty$, $j \in \{1, 2\}$.

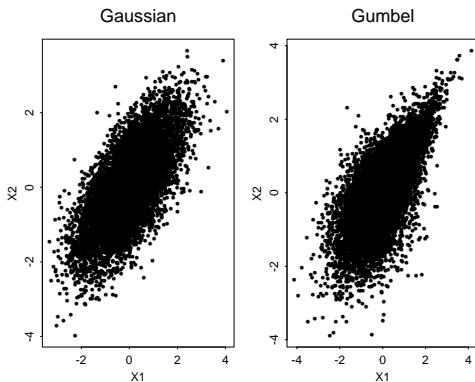
Note that ρ depends on the marginal distributions! In particular, second moments have to exist (not the case, e.g. for $X_1, X_2 \stackrel{\text{ind.}}{\sim} F(x) = 1 - x^{-3}!$)

- $|\rho| \leq 1$. Furthermore, $|\rho| = 1$ if and only if there are constants $a \in \mathbb{R} \setminus \{0\}, b \in \mathbb{R}$ with $X_2 = aX_1 + b$ a.s. with $a \geq 0$ if and only if $\rho = \pm 1$. This discards other strong functional dependence such as $X_2 = X_1^2$, for example.
- If X_1 and X_2 are independent, then $\rho = 0$. However, the converse is not true in general; see Example 7.17 below.
- ρ is invariant under strictly increasing linear transformations on $\text{ran } X_1 \times \text{ran } X_2$ but not invariant under strictly increasing functions in general. To see this, consider $(X_1, X_2) \sim N_2(\mathbf{0}, P)$ with $P_{12} = \rho$. Then $\rho(X_1, X_2) = \rho$, but $\rho(F_1(X_1), F_2(X_2)) = \frac{6}{\pi} \arcsin(\rho/2)$.

Correlation fallacies

Fallacy 1: F_1 , F_2 , and ρ uniquely determine F

This is **true for bivariate elliptical distributions**, **but wrong in general**. The following samples both have $N(0, 1)$ margins and correlation $\rho = 0.7$, yet come from different (copula) models:



Another example is this.

Example 7.17 (Uncorrelated \nRightarrow independent)

- Consider the two risks

$$X_1 = Z \quad (\text{Profit \& Loss Country A}),$$

$$X_2 = ZV \quad (\text{Profit \& Loss Country B}),$$

where V, Z are independent with $Z \sim N(0, 1)$ and $\mathbb{P}(V = -1) = \mathbb{P}(V = 1) = 1/2$. Then $X_2 \sim N(0, 1)$ and $\rho(X_1, X_2) = \text{cov}(X_1, X_2) = \mathbb{E}(X_1 X_2) \underset{\text{ind.}}{=} \mathbb{E}(V)\mathbb{E}(Z^2) = 0$, but X_1 and X_2 are not independent (in fact, V switches between counter- and comonotonicity).

- Consider $(X'_1, X'_2) \sim N_2(\mathbf{0}, I_2)$. Both (X'_1, X'_2) and (X_1, X_2) have $N(0, 1)$ margins and $\rho = 0$, but the copula of (X'_1, X'_2) is Π and the copula of (X_1, X_2) is the convex combination $C(\mathbf{u}) = \lambda M(\mathbf{u}) + (1 - \lambda)W(\mathbf{u})$ for $\lambda = 0.5$.

Fallacy 2: Given F_1, F_2 , any $\rho \in [-1, 1]$ is attainable

This is true for elliptically distributed (X_1, X_2) with $\mathbb{E}(R^2) < \infty$ (as then $\text{corr } \mathbf{X} = \mathbf{P}$), but wrong in general:

- If F_1 and F_2 are not of the same type (no linearity), $\rho(X_1, X_2) = 1$ is not attainable (recall that $|\rho| = 1$ if and only if there are constants $a \in \mathbb{R} \setminus \{0\}, b \in \mathbb{R}$ with $X_2 = aX_1 + b$ a.s.).
- What is the attainable range then? Hoeffding's identity

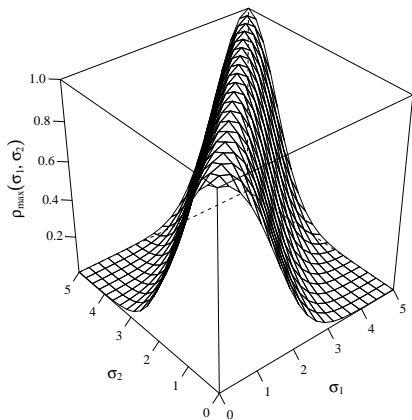
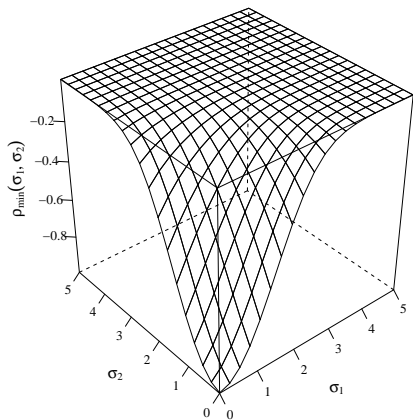
$$\text{cov}(X_1, X_2) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (C(F_1(x_1), F_2(x_2)) - F_1(x_1)F_2(x_2)) dx_1 dx_2.$$

implies bounds on attainable ρ :

$$\rho \in [\rho_{\min}, \rho_{\max}] \quad (\rho_{\min} \text{ is attained for } C = W, \rho_{\max} \text{ for } C = M).$$

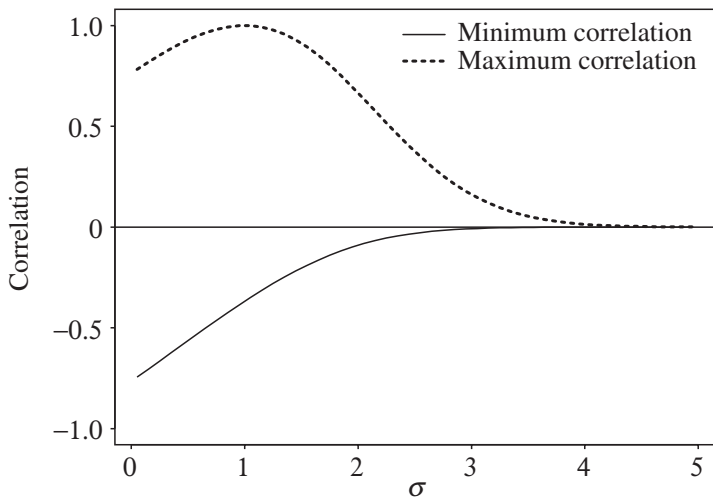
Example 7.18 (Bounds for a model with $\text{LN}(0, \sigma_j^2)$ margins)

Let $X_j \sim \text{LN}(0, \sigma_j^2)$, $j \in \{1, 2\}$. One can show that minimal (ρ_{\min} ; left) and maximal (ρ_{\max} ; right) correlations are given as follows.



For $\sigma_1^2 = 1$, $\sigma_2^2 = 16$ one has $\rho \in [-0.0003, 0.0137]!$

Specifically, let $X_1 \sim \text{LN}(0, 1)$ and $X_2 \sim \text{LN}(0, \sigma^2)$. Now let σ vary and plot ρ_{\min} and ρ_{\max} against σ :



Fallacy 3: ρ maximal (i.e. $C = M$) $\Rightarrow \text{VaR}_\alpha(X_1 + X_2)$ maximal

- This is true if (X_1, X_2) is elliptically distributed since the maximal $\rho = 1$ implies that X_1, X_2 are comonotone, so VaR_α is additive (by Proposition 7.15) and additivity provides the largest possible bound in this case as VaR_α is subadditive (by Proposition 6.24).
- Any superadditivity example $\text{VaR}_\alpha(X_1 + X_2) > \text{VaR}_\alpha(X_1) + \text{VaR}_\alpha(X_2)$ under comonotonicity (under comonotonicity, so maximal correlation qrm, the right-hand side is $\text{VaR}_\alpha(X_1 + X_2)$) serves as a counterexample; see Section 2.3.5.

7.2.3 Rank correlation

Rank correlation coefficients are...

- ... always defined;
- ... invariant under strictly increasing transformations of the random variables (hence only depend on the underlying copula).

Kendall's tau and Spearman's rho

Definition 7.19 (Kendall's tau)

Let $X_j \sim F_j$ with F_j continuous, $j \in \{1, 2\}$. Let (X'_1, X'_2) be an independent copy of (X_1, X_2) . *Kendall's tau* is defined by

$$\begin{aligned}\rho_\tau &= \mathbb{E}(\text{sign}((X_1 - X'_1)(X_2 - X'_2))) \\ &= \mathbb{P}((X_1 - X'_1)(X_2 - X'_2) > 0) - \mathbb{P}((X_1 - X'_1)(X_2 - X'_2) < 0),\end{aligned}$$

where $\text{sign}(x) = I_{(0, \infty)}(x) - I_{(-\infty, 0)}(x)$ (so -1 for $x < 0$, 0 for $x = 0$ and 1 for $x > 0$).

By definition, Kendall's tau is the probability of *concordance* minus the probability of *discordance*.

Proposition 7.20 (Formula for Kendall's tau)

Let $X_j \sim F_j$ with F_j continuous, $j \in \{1, 2\}$, and copula C . Then

$$\rho_\tau = 4 \int_0^1 \int_0^1 C(u_1, u_2) dC(u_1, u_2) - 1.$$

Proof. See the appendix. □

An estimator of ρ_τ is provided by the sample version of Kendall's tau

$$r_n^\tau = \frac{1}{\binom{n}{2}} \sum_{1 \leq i_1 < i_2 \leq n} \text{sign}((X_{i_1 1} - X_{i_2 1})(X_{i_1 2} - X_{i_2 2})). \quad (29)$$

Definition 7.21 (Spearman's rho)

Let $X_j \sim F_j$ with F_j continuous, $j \in \{1, 2\}$. *Spearman's rho* is defined by $\rho_S = \rho(F_1(X_1), F_2(X_2))$.

Proposition 7.22 (Formula for Spearman's rho)

Let $X_j \sim F_j$ with F_j continuous, $j \in \{1, 2\}$, and copula C . Then

$$\rho_S = 12 \int_0^1 \int_0^1 C(u_1, u_2) du_1 du_2 - 3.$$

Proof. By Hoeffding's identity, we have $\rho_S(X_1, X_2) = \rho(F_1(X_1), F_2(X_2)) = 12 \int_0^1 \int_0^1 (C(u_1, u_2) - u_1 u_2) du_1 du_2 = 12 \int_0^1 \int_0^1 C(u_1, u_2) du_1 du_2 - 3$. \square

- An estimator r_n^S is given by the sample correlation computed from componentwise (scaled) ranks (i.e. marginal empirical dfs) of the data.
- For $\kappa = \rho_\tau$ and $\kappa = \rho_S$, Embrechts et al. (2002) show that $\kappa = \pm 1$ if and only if X_1, X_2 are co-/countermonotonic.
- **Fallacy 1** (F_1, F_2, ρ uniquely determine F) is not solved by replacing ρ by rank correlation coefficients κ (it is easy to construct several copulas with the same Kendall's tau, e.g. via Archimedean copulas).

- **Fallacy 2** (For F_1, F_2 , any $\rho \in [-1, 1]$ is attainable) is solved. Take

$$F(x_1, x_2) = \lambda W(F_1(x_1), F_2(x_2)) + (1 - \lambda) M(F_1(x_1), F_2(x_2)).$$

This is a model with $\rho_S = \tau\rho_\tau = 1 - 2\lambda$ (choose λ as desired).

- **Fallacy 3** ($C = M$ implies $\text{VaR}_\alpha(X_1 + X_2)$ maximal) is also not solved by rank correlation coefficients $\kappa = 1$: Although $\kappa = 1$ corresponds to $C = M$, this copula does not necessarily provide the largest $\text{VaR}_\alpha(X_1 + X_2)$; see Fallacy 3 earlier.
- Also, in general, $\kappa = 0$ does not imply independence.
- Nevertheless, rank correlations are useful to summarize dependence, to parameterize copula families to make dependence comparable and for copula parameter calibration or estimation.

7.2.4 Coefficients of tail dependence

Goal: Measure **extremal dependence**, i.e. dependence in the **joint tails**.

Definition 7.23 (Tail dependence)

Let $X_j \sim F_j$, $j \in \{1, 2\}$, be continuously distributed random variables. Provided that the limits exist, the **lower tail-dependence coefficient** λ_l and **upper tail-dependence coefficient** λ_u of X_1 and X_2 are defined by

$$\lambda_l = \lim_{u \downarrow 0} \mathbb{P}(X_2 \leq F_2^{\leftarrow}(u) \mid X_1 \leq F_1^{\leftarrow}(u)),$$

$$\lambda_u = \lim_{u \uparrow 1} \mathbb{P}(X_2 > F_2^{\leftarrow}(u) \mid X_1 > F_1^{\leftarrow}(u)).$$

If $\lambda_l \in (0, 1]$ ($\lambda_u \in (0, 1]$), then (X_1, X_2) is **lower (upper) tail dependent**.
If $\lambda_l = 0$ ($\lambda_u = 0$), then (X_1, X_2) is **lower (upper) tail independent**.

As (conditional) probabilities, we clearly have $\lambda_l, \lambda_u \in [0, 1]$.

- Tail dependence is a copula property, since

$$\begin{aligned} \mathbb{P}(X_2 \leq F_2^{\leftarrow}(u) \mid X_1 \leq F_1^{\leftarrow}(u)) &= \frac{\mathbb{P}(X_1 \leq F_1^{\leftarrow}(u), X_2 \leq F_2^{\leftarrow}(u))}{\mathbb{P}(X_1 \leq F_1^{\leftarrow}(u))} \\ &= \frac{F(F_1^{\leftarrow}(u), F_2^{\leftarrow}(u))}{F_1(F_1^{\leftarrow}(u))} \stackrel{\text{Sklar}}{\underset{\text{(GI4)}}{=}} \frac{C(u, u)}{u}, \quad u \in (0, 1), \text{ so } \lambda_1 = \lim_{u \downarrow 0} \frac{C(u, u)}{u}. \end{aligned}$$

- If $u \mapsto C(u, u)$ is differentiable in a neighborhood of 0 and the limit exists, then $\lambda_1 = \lim_{u \downarrow 0} \frac{d}{du} C(u, u)$ (l'Hôpital's Rule).
- If C is totally differentiable in a neighborhood of 0 and the limit exists, then $\lambda_1 = \lim_{u \downarrow 0} (D_1 C(u, u) + D_2 C(u, u))$ (Chain Rule).
- If C is symmetric, $\lambda_1 = 2 \lim_{u \downarrow 0} D_1 C(u, u)$. By Theorem 7.13, $\lambda_1 = 2 \lim_{u \downarrow 0} \mathbb{P}(U_2 \leq u \mid U_1 = u)$ for $(U_1, U_2) \sim C$. Combined with any continuous df F and $(X_1, X_2) = (F^{\leftarrow}(U_1), F^{\leftarrow}(U_2))$, one has

$$\lambda_1 = 2 \lim_{x \downarrow -\infty} \mathbb{P}(X_2 \leq x \mid X_1 = x) \stackrel{\text{density}}{=} 2 \lim_{x \downarrow -\infty} \int_{-\infty}^x f_{X_2 \mid X_1=x}(x_2) dx_2. \quad (30)$$

- Similarly as above, for the upper tail-dependence coefficient,

$$\begin{aligned}\lambda_u &= \lim_{u \uparrow 1} \frac{1 - 2u + C(u, u)}{1 - u} = \lim_{u \downarrow 0} \frac{\hat{C}(u, u)}{u} \\ &= \lim_{u \uparrow 1} \frac{2(1 - u) - (1 - C(u, u))}{1 - u} = 2 - \lim_{u \uparrow 1} \frac{1 - C(u, u)}{1 - u}.\end{aligned}$$

- For all **radially symmetric copulas** (e.g. the bivariate C_P^{Ga} and $C_{\nu, P}^t$ copulas), we have $\lambda_l = \lambda_u =: \lambda$.
- For **Archimedean copulas with strict ψ** , a substitution and l'Hôpital's Rule show:

$$\begin{aligned}\lambda_l &= \lim_{u \downarrow 0} \frac{\psi(2\psi^{-1}(u))}{u} = \lim_{t \rightarrow \infty} \frac{\psi(2t)}{\psi(t)} = 2 \lim_{t \rightarrow \infty} \frac{\psi'(2t)}{\psi'(t)}, \\ \lambda_u &= 2 - \lim_{u \uparrow 1} \frac{1 - \psi(2\psi^{-1}(u))}{1 - u} = 2 - \lim_{t \downarrow 0} \frac{1 - \psi(2t)}{1 - \psi(t)} = 2 - 2 \lim_{t \downarrow 0} \frac{\psi'(2t)}{\psi'(t)}.\end{aligned}$$

Clayton: $\lambda_l = 2^{-1/\theta}$, $\lambda_u = 0$; **Gumbel:** $\lambda_l = 0$, $\lambda_u = 2 - 2^{1/\theta}$

7.3 Normal mixture copulas

... are the **copulas of multivariate normal** (mean-) **variance mixtures** $\mathbf{X} \stackrel{d}{=} \boldsymbol{\mu} + \sqrt{\mathbf{W}} \mathbf{A} \mathbf{Z}$ ($\mathbf{X} \stackrel{d}{=} \mathbf{m}(W) + \sqrt{W} \mathbf{A} \mathbf{Z}$); e.g. Gauss, t copulas.

7.3.1 Tail dependence

Coefficients of tail dependence

Let (X_1, X_2) be distributed according to a normal variance mixture and assume (w.l.o.g.) that $\boldsymbol{\mu} = (0, 0)$ and $\mathbf{A} \mathbf{A}' = \mathbf{P} = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}$. In this case, $F_1 = F_2$ and C is symmetric and radially symmetric. We thus obtain that

$$\lambda \stackrel{\text{radial}}{=} \lambda \stackrel{\text{symm.}}{=} \lambda \stackrel{\text{symm.}}{=} \lim_{x \downarrow -\infty} \mathbb{P}(X_2 \leq x \mid X_1 = x).$$

(30)

Example 7.24 (λ for the Gauss and t copula)

- Considering the bivariate $N(\mathbf{0}, \mathbf{P})$ density, one can show (via $f_{X_2|X_1}(x_2 \mid x_1) = \frac{f_{X_1, X_2}(x_1, x_2)}{f_{X_1}(x_1)}$) that $X_2 \mid X_1 = x \sim N(\rho x, 1 - \rho^2)$. This implies that

$\lambda = 2 \lim_{x \downarrow -\infty} \mathbb{P}(X_2 \leq x \mid X_1 = x) = 2 \lim_{x \downarrow -\infty} \Phi\left(\frac{x(1-\rho)}{\sqrt{1-\rho^2}}\right) = I_{\{\rho=1\}}$
 (essentially **no tail dependence**).

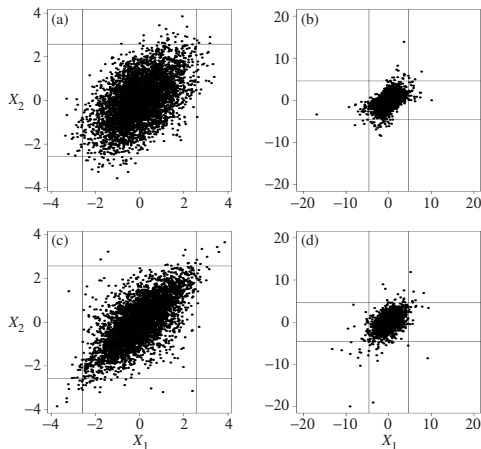
- For $C_{\nu, P}^t$, one can show that $X_2 \mid X_1 = x \sim t_{\nu+1}\left(\rho x, \frac{(1-\rho^2)(\nu+x^2)}{\nu+1}\right)$ and thus $\mathbb{P}(X_2 \leq x \mid X_1 = x) = t_{\nu+1}\left(\frac{x-\rho x}{\sqrt{\frac{(1-\rho^2)(\nu+x^2)}{\nu+1}}}\right)$. Hence

$$\lambda = 2t_{\nu+1}\left(-\sqrt{\frac{(\nu+1)(1-\rho)}{1+\rho}}\right) \quad (\text{tail dependence}).$$

ν	$\rho = -0.5$	$\rho = 0$	$\rho = 0.5$	$\rho = 0.9$	$\rho = 1$
∞	0	0	0	0	1
10	0.00	0.01	0.08	0.46	1
4	0.01	0.08	0.25	0.63	1
2	0.06	0.18	0.39	0.72	1

What drives tail dependence of normal variance mixtures is W . If W has a power tail, we get tail dependence, otherwise not.

Joint quantile exceedance probabilities



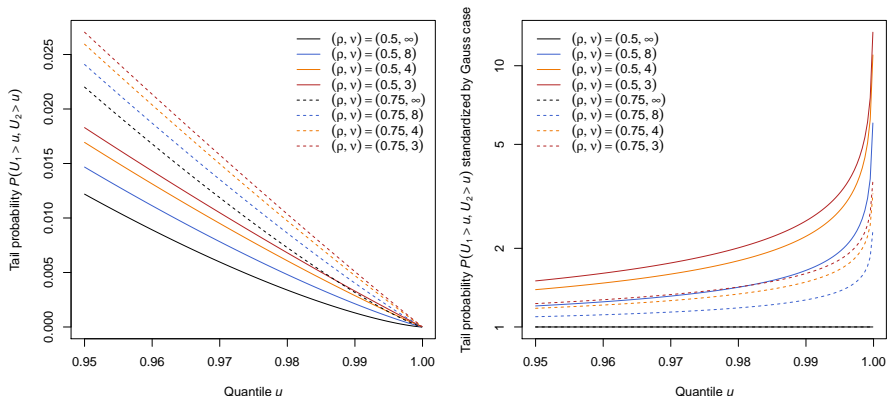
5000 samples from

- (a) $N_2(\mathbf{0}, P = (\frac{1}{\rho} \begin{pmatrix} \rho & 1 \\ 1 & \rho \end{pmatrix}))$, $\rho = 0.5$;
- (b) C_{ρ}^{Ga} with t_4 margins (same dependence as in (a));
- (c) $C_{4,\rho}^t$ with $N(0, 1)$ margins;
- (d) $t_2(4, \mathbf{0}, P)$ (same dependence as in (c)).

Lines denote 0.005- and 0.995-quantiles.

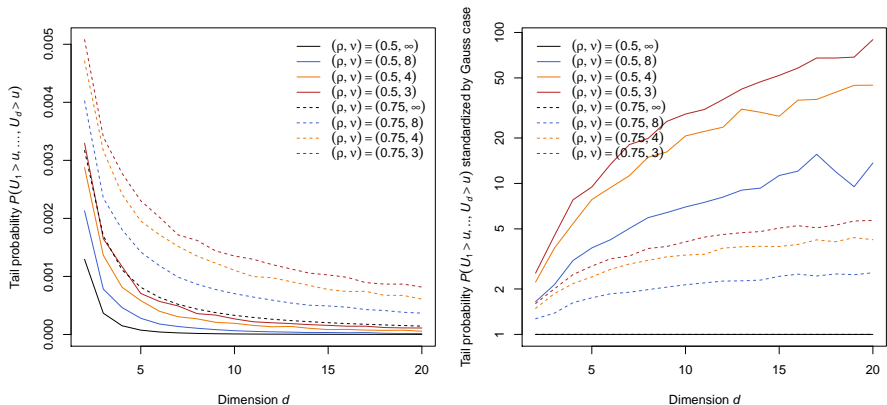
Note the different number of points in the bivariate tails (all models have the same Kendall's tau!)

Joint tail probabilities $\mathbb{P}(U_1 > u, U_2 > u)$ for $d = 2$



- **Left:** The higher ρ or the smaller ν , the larger $\mathbb{P}(U_1 > u, U_2 > u)$.
- **Right:** $u \mapsto \frac{\mathbb{P}(U_1 > u, U_2 > u)}{\mathbb{P}(V_1 > u, V_2 > u)} \underset{\text{symm.}}{\stackrel{\text{radial}}{=}} \frac{C_{\nu, \rho}^t(u, u)}{C_{\rho}^{\text{Ga}}(u, u)}$

Joint tail probabilities $\mathbb{P}(U_1 > u, \dots, U_d > u)$ for $u = 0.99$



- Homogeneous P (off-diagonal entry ρ). Note the MC randomness.
- **Left:** Clear, less mass in corners in higher dimensions.
- **Right:** $d \mapsto \frac{\mathbb{P}(U_1 > u, \dots, U_d > u)}{\mathbb{P}(V_1 > u, \dots, V_d > u)} \underset{\text{symm.}}{\text{radial}} \frac{C_{\nu, \rho}^t(u, \dots, u)}{C_{\rho}^{\text{Ga}}(u, \dots, u)}$ for $u = 0.99$.

Example 7.25 (Interpretation of joint tail probabilities)

- Consider 5 daily returns $\mathbf{X} = (X_1, \dots, X_5)$ with pairwise correlations (all) $\rho = 0.5$. However, we are unsure about the best joint model.
- If the copula of \mathbf{X} is $C_{\rho=0.5}^{\text{Ga}}$, the probability that on any day all 5 returns lie below their $u = 0.01$ quantiles is

$$\mathbb{P}(X_1 \leq F_1^{\leftarrow}(u), \dots, X_5 \leq F_5^{\leftarrow}(u)) = \mathbb{P}(U_1 \leq u, \dots, U_5 \leq u) \\ \approx_{\text{MC error}} 7.48 \times 10^{-5}.$$

In the long run such an event will happen once every $1/7.48 \times 10^{-5} \approx 13\,369$ trading days on average (\approx once every 51.4 years; assuming 260 trading days in a year).

- If the copula of \mathbf{X} is $C_{\nu=4, \rho=0.5}^t$, however, such an event will happen approximately 7.68 times more often, i.e. \approx once every 6.7 years. This gets worse the larger d !

7.3.2 Rank correlations

Proposition 7.26 (Spearman's rho for normal variance mixtures)

Let $\mathbf{X} \sim M_2(\mathbf{0}, P, \hat{F}_W)$ with $\mathbb{P}(\mathbf{X} = \mathbf{0}) = 0$, $\rho = P_{12}$. Then

$$\rho_S = \frac{6}{\pi} \mathbb{E} \left(\arcsin \frac{W\rho}{\sqrt{(W + \tilde{W})(W + \bar{W})}} \right),$$

for $W, \tilde{W}, \bar{W} \stackrel{\text{ind.}}{\sim} F_W$ with Laplace–Stieltjes transform \hat{F}_W . For Gauss copulas, $\rho_S = \frac{6}{\pi} \arcsin(\frac{\rho}{2})$.

Proof. See the appendix. □

Proposition 7.27 (Kendall's tau for elliptical distributions)

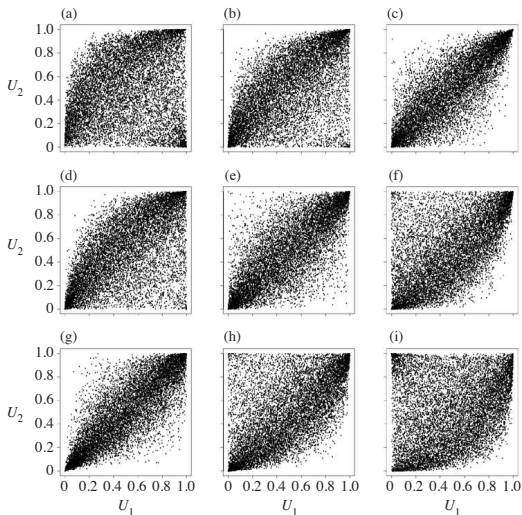
Let $\mathbf{X} \sim E_2(\mathbf{0}, P, \psi)$ with $\mathbb{P}(\mathbf{X} = \mathbf{0}) = 0$, $\rho = P_{12}$. Then $\rho_\tau = \frac{2}{\pi} \arcsin \rho$.

Proof. See the appendix. □

7.3.3 Skewed normal mixture copulas

- *Skewed normal mixture copulas* are the copulas of normal mixture distributions which are not elliptical, e.g. the *skewed t copula* $C_{\nu, P, \gamma}^t$ is the copula of a generalized hyperbolic distribution; see McNeil et al. (2015, Sections 6.2.3 and 7.3.3) for more details.
- It can be sampled as other implicit copulas; see Algorithm 7.9 (the evaluation of the margins requires numerical integration of a skewed t density).
- The main advantage of such a copula over $C_{\nu, P}^t$ is its radial asymmetry (e.g. for modelling $\lambda_l \neq \lambda_u$)

10 000 samples from $C_{\nu=5, \rho=0.8, \gamma=0.8(I_{\{i<2\}}-I_{\{i>2\}}, I_{\{j>2\}}-I_{\{j<2\}})}$:



(a) $\gamma = (0.8, -0.8)$

(b) $\gamma = (0.8, 0)$

(c) $\gamma = (0.8, 0.8)$

(d) $\gamma = (0, -0.8)$

(e) $\gamma = (0, 0)$

(f) $\gamma = (0, 0.8)$

(g) $\gamma = (-0.8, -0.8)$

(h) $\gamma = (-0.8, 0)$

(i) $\gamma = (-0.8, 0.8)$

7.3.4 Grouped normal mixture copulas

- *Grouped normal mixture copulas* are copulas which attach together a set of normal mixture copulas, e.g. a *grouped t copula* is the copula of

$$\mathbf{X} = (\sqrt{W_1}Y_1, \dots, \sqrt{W_1}Y_{s_1}, \dots, \sqrt{W_S}Y_{s_1+\dots+s_{S-1}+1}, \dots, \sqrt{W_S}Y_d)$$

for $(W_1, \dots, W_S) \sim M(\text{IG}(\frac{\nu_1}{2}, \frac{\nu_1}{2}), \dots, \text{IG}(\frac{\nu_S}{2}, \frac{\nu_S}{2}))$ and $\mathbf{Y} \sim N_d(\mathbf{0}, P)$ (so $\mathbf{Y} \stackrel{d}{=} A\mathbf{Z}$ as before); see Demarta and McNeil (2005) for details.

- Clearly, the marginals are t distributed, hence

$$\mathbf{U} = (t_{\nu_1}(X_1), \dots, t_{\nu_1}(X_{s_1}), \dots, t_{\nu_S}(X_{s_1+\dots+s_{S-1}+1}), \dots, t_{\nu_S}(X_d))$$

follows a *grouped t copula*. This is straightforward to simulate.

- It can be fitted with pairwise inversion of Kendall's tau.
- If $S = d$, grouped t copulas are also known as *generalized t copulas*; see Luo and Shevchenko (2010).

7.4 Archimedean copulas

Recall that an (Archimedean) generator ψ is a function $\psi : [0, \infty) \rightarrow [0, 1]$ which is \downarrow on $[0, \inf\{t : \psi(t) = 0\}]$ and satisfies $\psi(0) = 1$, $\psi(\infty) = \lim_{t \rightarrow \infty} \psi(t) = 0$; the set of all generators is denoted by Ψ .

7.4.1 Bivariate Archimedean copulas

Theorem 7.28 (Bivariate Archimedean copulas)

For $\psi \in \Psi$, $C(u_1, u_2) = \psi(\psi^{-1}(u_1) + \psi^{-1}(u_2))$ is a copula if and only if ψ is convex.

- For a strict and twice-continuously differentiable ψ , one can show that

$$\rho_\tau = 1 - 4 \int_0^\infty t(\psi'(t))^2 dt = 1 + 4 \int_0^1 \frac{\psi^{-1}(t)}{(\psi^{-1}(t))'} dt.$$

- If ψ is strict, $\lambda_l = 2 \lim_{t \rightarrow \infty} \frac{\psi'(2t)}{\psi'(t)}$ and $\lambda_u = 2 - 2 \lim_{t \downarrow 0} \frac{\psi'(2t)}{\psi'(t)}$.

- The most widely used one-parameter Archimedean copulas are:

Family	θ	$\psi(t)$	$V \sim F = \mathcal{LS}^{-1}(\psi)$
A	$[0, 1)$	$(1 - \theta)/(\exp(t) - \theta)$	Geo($1 - \theta$)
C	$(0, \infty)$	$(1 + t)^{-1/\theta}$	$\Gamma(1/\theta, 1)$
F	$(0, \infty)$	$-\log(1 - (1 - e^{-\theta}) \exp(-t))/\theta$	Log($1 - e^{-\theta}$)
G	$[1, \infty)$	$\exp(-t^{1/\theta})$	$S(1/\theta, 1, \cos^{\theta}(\pi/(2\theta)), I_{\{\theta=1\}}; 1)$
J	$[1, \infty)$	$1 - (1 - \exp(-t))^{1/\theta}$	Sibuya($1/\theta$)

Family	ρ_{τ}	λ_l	λ_u
A	$1 - 2(\theta + (1 - \theta)^2 \log(1 - \theta))/(3\theta^2)$	0	0
C	$\theta/(\theta + 2)$	$2^{-1/\theta}$	0
F	$1 + 4(D_1(\theta) - 1)/\theta$	0	0
G	$(\theta - 1)/\theta$	0	$2 - 2^{1/\theta}$
J	$1 - 4 \sum_{k=1}^{\infty} 1/(k(\theta k + 2)(\theta(k - 1) + 2))$	0	$2 - 2^{1/\theta}$

7.4.2 Multivariate Archimedean copulas

ψ is *completely monotone (c.m.)* if $(-1)^k \psi^{(k)}(t) \geq 0$ for all $t \in (0, \infty)$ and all $k \in \mathbb{N}_0$. The set of all c.m. generators is denoted by Ψ_∞ .

Theorem 7.29 (Kimberling (1974))

If $\psi \in \Psi$, $C(\mathbf{u}) = \psi\left(\sum_{j=1}^d \psi^{-1}(u_j)\right)$ is a copula $\forall d$ if and only if $\psi \in \Psi_\infty$.

Bernstein's Theorem characterizes all $\psi \in \Psi_\infty$.

Theorem 7.30 (Bernstein (1928))

$\psi(0) = 1$, ψ c.m. if and only if $\psi(t) = \mathbb{E}(\exp(-tV))$ for $V \sim G$ with $V \geq 0$ and $G(0) = 0$.

We thus use the notation $\psi = \hat{G}$.

Proposition 7.31 (Stochastic representation, related properties)

Let $\psi \in \Psi_\infty$ with $V \sim G$ such that $\hat{G} = \psi$ and let $E_1, \dots, E_d \stackrel{\text{ind.}}{\sim} \text{Exp}(1)$ be independent of V . Then

- 1) The survival copula of $\mathbf{X} = (\frac{E_1}{V}, \dots, \frac{E_d}{V})$ is Archimedean (with ψ).
- 2) $\mathbf{U} = (\psi(X_1), \dots, \psi(X_d)) \sim C$ and the U_j 's are conditionally independent given V with $\mathbb{P}(U_j \leq u \mid V = v) = \exp(-v\psi^{-1}(u))$.

Proof.

- 1) The joint survival function of \mathbf{X} is given by

$$\begin{aligned}\bar{F}(\mathbf{x}) &= \mathbb{P}(X_j > x_j \ \forall j) = \int_0^\infty \mathbb{P}(E_j/V > x_j \ \forall j \mid V = v) dG(v) \\ &= \int_0^\infty \mathbb{P}(E_j > vx_j \ \forall j) dG(v) = \int_0^\infty \prod_{j=1}^d \exp(-vx_j) dG(v) \\ &= \int_0^\infty \exp\left(-v \sum_{j=1}^d x_j\right) dG(v) = \psi\left(\sum_{j=1}^d x_j\right).\end{aligned}$$

The j th marginal survival function is thus (set $x_k = 0 \ \forall k \neq j$)
 $\bar{F}_j(x_j) = \mathbb{P}(X_j > x_j) = \psi(x_j)$ (\downarrow and continuous) and therefore
 $\hat{C}(\mathbf{u}) = \bar{F}(\bar{F}_1^{\leftarrow}(u_1), \dots, \bar{F}_d^{\leftarrow}(u_d)) = \psi(\sum_{j=1}^d \psi^{-1}(u_j))$.

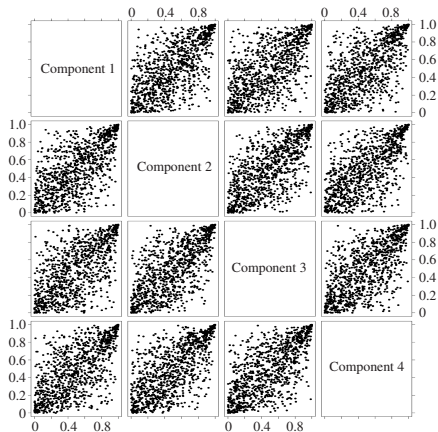
- 2) $\mathbb{P}(\mathbf{U} \leq \mathbf{u}) = \mathbb{P}(X_j > \psi^{-1}(u_j) \ \forall j) \stackrel{1)}{=} \psi(\sum_{j=1}^d \psi^{-1}(u_j))$. Conditional independence is clear by construction and $\mathbb{P}(U_j \leq u \mid V = v) = \mathbb{P}(X_j > \psi^{-1}(u) \mid V = v) = \mathbb{P}(E_j > v\psi^{-1}(u)) = \exp(-v\psi^{-1}(u))$. \square

We call all Archimedean copulas with $\psi \in \Psi_\infty$ *LT-Archimedean copulas*.

Algorithm 7.32 (Marshall and Olkin (1988))

- 1) Sample $V \sim G$ (df corresponding to ψ).
- 2) Sample $E_1, \dots, E_d \stackrel{\text{ind.}}{\sim} \text{Exp}(1)$ independently of V .
- 3) Return $\mathbf{U} = (\psi(E_1/V), \dots, \psi(E_d/V))$ (conditional independence).

1000 samples of a 4-dim. Gumbel copula ($\rho_\tau = 0.5$; $\lambda_u \approx 0.5858$)



- For fixed d , c.m. can be relaxed to d -monotonicity; see McNeil and Nešlehová (2009).
- Various non-exchangeable extensions to Archimedean copulas exist.

7.5 Fitting copulas to data

- Let $\mathbf{X}, \mathbf{X}_1, \dots, \mathbf{X}_n$ be independent random vectors with df F , continuous margins F_1, \dots, F_d and copula C . We assume we have data $\mathbf{x}_1, \dots, \mathbf{x}_n$, interpreted as realizations of $\mathbf{X}_1, \dots, \mathbf{X}_n$; in what follows we work with the latter.
- Assume
 - ▶ $F_j = F_j(\cdot; \theta_{0,j})$ for some $\theta_{0,j} \in \Theta_j$, $j \in \{1, \dots, d\}$;
($F_j(\cdot; \theta_j)$ continuous $\forall \theta_j \in \Theta_j$, $j \in \{1, \dots, d\}$)
 - ▶ $C = C(\cdot; \theta_{0,C})$ for some $\theta_{0,C} \in \Theta_C$.

Thus F has the true but unknown parameter vector $\theta_0 = (\theta'_{0,C}, \theta'_{0,1}, \dots, \theta'_{0,d})'$ to be estimated.

- Here, we focus particularly on $\theta_{0,C}$. Whenever necessary, we assume that the margins F_1, \dots, F_d and the copula C are absolutely continuous with corresponding densities f_1, \dots, f_d and c , respectively.

- We assume the chosen copula to be appropriate (w.r.t. symmetry, tail dependence etc.).

7.5.1 Method-of-moments using rank correlation

- We focus on one-parameter copulas here, i.e. $\theta_{0,C} = \theta_0$.
- For $d = 2$, Genest and Rivest (1993) suggested estimating $\theta_{0,C}$ by solving $\rho_\tau(\theta_C) = r_n^\tau$ w.r.t. θ_C , i.e.

$$\hat{\theta}_{n,C}^{\text{IKTE}} = \rho_\tau^{-1}(r_n^\tau), \quad (\text{inversion of Kendall's tau estimator (IKTE)})$$

where $\rho_\tau(\cdot)$ denotes Kendall's tau as a function in θ and r_n^τ is the sample version of Kendall's tau (computed via (29) from $\mathbf{X}_1, \dots, \mathbf{X}_n$ or pseudo-observations U_1, \dots, U_n ; see later).

- The standardized dispersion matrix P for elliptical copulas can be estimated via *pairwise inversion of Kendall's tau*; see McNeil et al. (2015, Example 7.56). If $r_{n,j_1j_2}^\tau$ denotes the sample version of Kendall's tau for data pair (j_1, j_2) , then $\hat{P}_{n,j_1j_2}^{\text{IKTE}} = \sin(\frac{\pi}{2} r_{n,j_1j_2}^\tau)$; see Proposition 7.27.

For obtaining a proper correlation matrix P (positive semi-definite), see Higham (2002).

- ▶ For Gauss copulas, it is preferable to use Spearman's rho based on

$$\rho_S \stackrel{\text{Prop. 7.26}}{=} \frac{6}{\pi} \arcsin \frac{\rho}{2} \approx \rho.$$

The latter approximation error is comparably small, so that the matrix of pairwise sample versions of Spearman's rho is an estimator for P .

- ▶ For t copulas, \hat{P}_n^{IKTE} can be used to estimate P and then ν can be estimated via its MLE based on \hat{P}_n^{IKTE} .

7.5.2 Forming a pseudo-sample from the copula

- X_1, \dots, X_n (as good as) never has $U(0, 1)$ margins. For applying the “copula approach” we thus need *pseudo-observations* from C .
- In general, we take $\hat{U}_i = (\hat{U}_{i1}, \dots, \hat{U}_{id}) = (\hat{F}_1(X_{i1}), \dots, \hat{F}_d(X_{id}))$, $i \in \{1, \dots, n\}$, where \hat{F}_j denotes an estimator of F_j ; see Lemma 7.6. Note

that $\hat{U}_1, \dots, \hat{U}_n$ are typically neither independent (even if X_1, \dots, X_n are) nor perfectly $U(0, 1)$.

■ Possible choices for \hat{F}_j :

- 1) Non-parametric estimators with scaled empirical dfs (to avoid density evaluation on the boundary of $[0, 1]^d$), so

$$\hat{U}_{ij} = \frac{n}{n+1} \hat{F}_{n,j}(X_{ij}) = \frac{R_{ij}}{n+1}, \quad (31)$$

where R_{ij} denotes the rank of X_{ij} among all X_{1j}, \dots, X_{nj} .

- 2) Parametric estimators (such as Student t , Pareto, etc.; typically if n is small). In this case, one often still uses (31) for estimating $\theta_{0,C}$ (to keep the error due to misspecification of the margins small).
- 3) EVT-based. Bodies are modelled empirically; tails semiparametrically via GPD.

7.5.3 Maximum likelihood estimation

The (classical) maximum likelihood estimator

- By Sklar's Theorem, the density of F is given by

$$f(x; \theta_0) = c(F_1(x_1; \theta_{0,1}), \dots, F_d(x_d; \theta_{0,d}); \theta_{0,C}) \prod_{j=1}^d f_j(x_j; \theta_{0,j}).$$

- The log-likelihood based on $\mathbf{X}_1, \dots, \mathbf{X}_n$ is thus

$$\begin{aligned} \ell(\theta; \mathbf{X}_1, \dots, \mathbf{X}_n) &= \sum_{i=1}^n \ell(\theta; \mathbf{X}_i) \\ &= \sum_{i=1}^n \ell_C(\theta_C; F_1(X_{i1}; \theta_1), \dots, F_d(X_{id}; \theta_d)) + \sum_{i=1}^n \sum_{j=1}^d \ell_j(\theta_j; X_{ij}), \end{aligned}$$

where

$$\ell_C(\theta_C; u_1, \dots, u_d) = \log c(u_1, \dots, u_d; \theta_C)$$

$$\ell_j(\theta_j; x) = \log f_j(x; \theta_j), \quad j \in \{1, \dots, d\}.$$

- The *maximum likelihood estimator (MLE)* of θ_0 is

$$\hat{\theta}_n^{\text{MLE}} = \underset{\theta \in \Theta}{\operatorname{argsup}} \ell(\theta; \mathbf{X}_1, \dots, \mathbf{X}_n).$$

This optimization is typically done by numerical means. Note that this can be quite demanding, especially in high dimensions.

The inference functions for margins estimator

- Joe and Xu (1996) suggested the *two-step estimation approach*:

Step 1: For $j \in \{1, \dots, d\}$, estimate $\theta_{0,j}$ by its MLE $\hat{\theta}_{n,j}^{\text{MLE}}$.

Step 2: Estimate $\theta_{0,C}$ by

$$\hat{\theta}_{n,C}^{\text{IFME}} = \underset{\theta_C \in \Theta_C}{\operatorname{argsup}} \ell(\theta_C, \hat{\theta}_{n,1}^{\text{MLE}}, \dots, \hat{\theta}_{n,d}^{\text{MLE}}; \mathbf{X}_1, \dots, \mathbf{X}_n).$$

The *inference functions for margins estimator (IFME)* of θ_0 is thus

$$\hat{\theta}_n^{\text{IFME}} = (\hat{\theta}_{n,C}^{\text{IFME}}, \hat{\theta}_{n,1}^{\text{MLE}}, \dots, \hat{\theta}_{n,d}^{\text{MLE}})$$

- This is typically much easier to compute than $\hat{\theta}_n^{\text{MLE}}$ while providing good results; see Joe and Xu (1996) or Kim et al. (2007).
- $\hat{\theta}_n^{\text{IFME}}$ can also be used as initial value for computing $\hat{\theta}_n^{\text{MLE}}$.
- In terms of likelihood equations, $\hat{\theta}_n^{\text{IFME}}$ compares to $\hat{\theta}_n^{\text{MLE}}$ as follows:

$$\begin{aligned}\hat{\theta}_n^{\text{MLE}} \text{ solves } \left(\frac{\partial}{\partial \theta_C} \ell, \frac{\partial}{\partial \theta_1} \ell, \dots, \frac{\partial}{\partial \theta_d} \ell \right) &= \mathbf{0}, \\ \hat{\theta}_n^{\text{IFME}} \text{ solves } \left(\frac{\partial}{\partial \theta_C} \ell, \frac{\partial}{\partial \theta_1} \ell_1, \dots, \frac{\partial}{\partial \theta_d} \ell_d \right) &= \mathbf{0},\end{aligned}$$

where

$$\begin{aligned}\ell &= \ell(\boldsymbol{\theta}; \mathbf{X}_1, \dots, \mathbf{X}_n), \\ \ell_j &= \ell_j(\boldsymbol{\theta}_j; X_{1j}, \dots, X_{nj}) = \sum_{i=1}^n \ell_j(\boldsymbol{\theta}_j; X_{ij}).\end{aligned}$$

Example 7.33 (A computationally convincing example)

Suppose $X_j \sim N(\mu_j, \sigma_j^2)$, $j \in \{1, \dots, d\}$, for $d = 100$, and C has (just) one parameter.

- MLE requires to solve a 201-dimensional optimization problem.
- IFME only requires 100 optimizations in two dimensions and 1 one-dimensional optimization.

If the marginals are estimated parametrically one often still uses the pseudo-observations built from the marginal empirical dfs to estimate $\theta_{0,C}$ (see MPLE below) in order to avoid misspecification of the margins (if n is sufficiently large).

The maximum pseudo-likelihood estimator

- The *maximum pseudo-likelihood estimator (MPLE)*, introduced by Genest et al. (1995), works similarly to $\hat{\theta}_n^{\text{IFME}}$, but estimates the margins non-parametrically:

Step 1: Compute rank-based pseudo-observations $\hat{U}_1, \dots, \hat{U}_n$.

Step 2: Estimate $\theta_{0,C}$ by

$$\hat{\theta}_{n,C}^{\text{MPLE}} = \underset{\theta_C \in \Theta_C}{\operatorname{argsup}} \sum_{i=1}^n \ell_C(\theta_C; \hat{U}_{i1}, \dots, \hat{U}_{id}) = \underset{\theta_C \in \Theta_C}{\operatorname{argsup}} \sum_{i=1}^n \log c(\hat{U}_i; \theta_C).$$

- Genest and Werker (2002) show that $\hat{\theta}_{n,C}^{\text{MPLE}}$ is not asymptotically efficient in general.
- Kim et al. (2007) compare $\hat{\theta}_n^{\text{MLE}}$, $\hat{\theta}_n^{\text{IFME}}$, and $\hat{\theta}_{n,C}^{\text{MPLE}}$ in a simulation study ($d = 2$ only!) and argue in favor of $\hat{\theta}_{n,C}^{\text{MPLE}}$ overall, especially w.r.t. robustness against misspecification of the margins; but see Embrechts and Hofert (2013b) for $d \gg 2$.

Example 7.34 (Fitting the Gauss copula)

- The (copula-related) log-likelihood ℓ_C is

$$\ell_C(P; \hat{U}_1, \dots, \hat{U}_n) = \sum_{i=1}^n \ell_C(P; \hat{U}_i) \stackrel{\text{Eq. (27)}}{=} \sum_{i=1}^n \log c_P^{\text{Ga}}(\hat{U}_i).$$

For maximization over all correlation matrices P , we can use the Cholesky factor A as reparameterization and maximize over all lower triangular matrices A with 1s on the diagonal; still this is $\mathcal{O}(d^2)$.

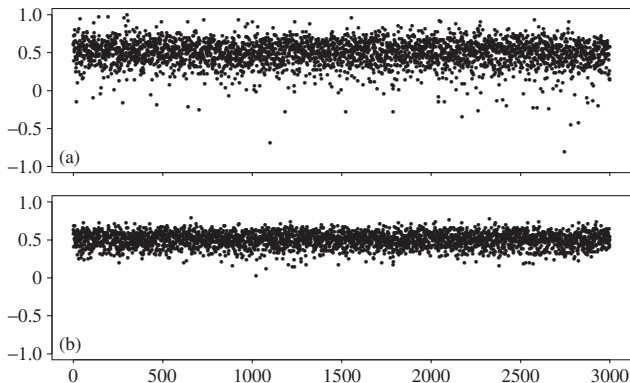
- Alternatively, use pairwise inversion of Spearman's rho or Kendall's tau.

Example 7.35 (Fitting the t copula)

- For small d , maximize the likelihood over all correlation matrices (as for the Gauss copula case) and the d.o.f. ν .
- For moderate/larger d , do:
 - 1) Estimate P via pairwise inversion of Kendall's tau (see above).
 - 2) Plug \hat{P} into the likelihood and maximize it w.r.t. ν to obtain $\hat{\nu}_n$.

Example 7.36 (Correlation estimation for heavy-tailed data)

Consider $n = 3000$ realizations of independent samples of size 90 from $t_2(3, \mathbf{0}, (\frac{1}{0.5} \ 0.5))$ (\Rightarrow linear correlation $\rho = 0.5$). Shall we estimate ρ via the sample correlation (estimates are shown in (a)) or via inversion of Kendall's tau (shown in (b))? The variance of the latter is smaller:



Estimation is only one side of the coin. The other is *goodness-of-fit* (i.e. to find out whether our estimated model indeed represents the given data well) and *model selection* (i.e. to decide which model is best among all adequate fitted models). Goodness-of-fit can be (computationally) challenging, particularly for large d . See the appendix for a graphical approach.

7.6 A copulas-based proof of subadditivity of ES

Proposition 7.37 (Subadditivity of ES)

$$\text{ES}_\alpha(L) = \frac{\sup_{\{\tilde{Y} \sim B(1, 1-\alpha)\}} \mathbb{E}(L\tilde{Y})}{1 - \alpha}, \text{ which, trivially, is subadditive.}$$

Proof. Let $L = F_L^\leftarrow(U)$ for $U \sim U(0, 1)$ and $Y = I_{\{U > \alpha\}} \sim B(1, 1 - \alpha)$. Then $\text{ES}_\alpha(L) = \frac{1}{1-\alpha} \int_\alpha^1 F_L^\leftarrow(u) du = \frac{1}{1-\alpha} \int_0^1 F_L^\leftarrow(u) I_{\{u > \alpha\}} \cdot 1 du = \frac{1}{1-\alpha} \mathbb{E}(F_L^\leftarrow(U) I_{\{U > \alpha\}}) = \frac{1}{1-\alpha} \mathbb{E}(LY)$. Note that L and Y are comontone, so that for any other $\tilde{Y} \sim B(1, 1 - \alpha)$, Hoeffding's identity implies that $\mathbb{E}(L\tilde{Y}) \leq \mathbb{E}(LY)$. Hence $\text{ES}_\alpha(L) = \frac{\sup_{\{\tilde{Y} \sim B(1, 1-\alpha)\}} \mathbb{E}(L\tilde{Y})}{(1 - \alpha)}$. From this representation, ES_α is easily seen to be subadditive. \square

This is also the shortest proof according to Embrechts and Wang (2015). An elementary proof is given in the appendix.