7 Copulas and dependence

7.1 Copulas

- We now look more closely at modeling the dependence among the components of a random vector $\boldsymbol{X} \sim H$ (risk-factor changes).
- **In short:** H "$=$" 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 H, 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 $\mathrm{U}[0,1]$ margins.

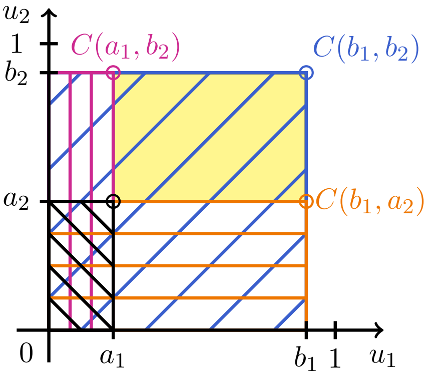
Characterization

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

1) C is *grounded*, that is,
   $C(u_1, \dots, u_d) = 0$ if $u_j = 0$ for at least one $j \in \{1, \dots, d\}$.
2) C has standard *uniform* univariate *margins*, that is,
   $C(1, \dots, 1, u_j, 1, \dots, 1) = u_j$ for all $u_j \in [0,1]$ and $j \in \{1, \dots, d\}$.
3) C is d*-increasing*, that is, for all $\boldsymbol{a}, \boldsymbol{b} \in [0,1]^d$, $\boldsymbol{a} \le \boldsymbol{b}$,
   $$\Delta_{(\boldsymbol{a},\boldsymbol{b}]} C = \sum_{\boldsymbol{i} \in \{0,1\}^d} (-1)^{\sum_{j=1}^d i_j} C(a_1^{i_1} b_1^{1-i_1}, \dots, a_d^{i_d} b_d^{1-i_d}) \ge 0.$$
   Equivalently (if existent): density $c(\boldsymbol{u}) \ge 0$ for all $\boldsymbol{u} \in (0,1)^d$.

2-increasingness explained in a picture:

$$\Delta_{(\boldsymbol{a},\boldsymbol{b}]}C = C(b_1,b_2) - C(b_1,a_2) - C(a_1,b_2) + C(a_1,a_2)$$
$$= \mathbb{P}(\boldsymbol{U} \in (\boldsymbol{a},\boldsymbol{b}]) \overset{!}{\geq} 0$$



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

Lemma 7.2

For all $\boldsymbol{a}, \boldsymbol{b} \in [0,1]^d$,

$$|C(\boldsymbol{b}) - C(\boldsymbol{a})| \le \sum_{j=1}^{d} |b_j - a_j| \tag{34}$$

In particular, copulas are uniformly equi-continuous.

Proof. Exanding $C(\boldsymbol{b}) - C(\boldsymbol{a})$ in a telescoping sum and using the triangle inequality leads to

$$\begin{aligned}|C(\boldsymbol{b}) - C(\boldsymbol{a})| \le \sum_{j=1}^{d} |C(b_1, \dots, b_{j-1}, b_j, a_{j+1}, \dots, a_d) \\ - C(b_1, \dots, b_{j-1}, a_j, a_{j+1}, \dots, a_d)|\end{aligned}$$

W.l.o.g. let $\boldsymbol{a} \le \boldsymbol{b}$. By d-increasingness, $C \nearrow$ in each component, so omit $|\cdot|$. Since, again by d-increasingness, the jth summand is $\nearrow$ in each component $\neq j$, let $b_1, \dots, b_{j-1}, a_{j+1}, \dots, a_d \nearrow 1$ to obtain the upper bound $\sum_{j=1}^{d} C(1, \dots, 1, b_j, 1, \dots, 1) - C(1, \dots, 1, a_j, 1, \dots, 1) = b_j - a_j$ for summand j. $\square$

A first "warm-up" example:

Let C_1, C_2 be copulas. Then $C(\boldsymbol{u}) = \lambda C_1(\boldsymbol{u}) + (1-\lambda)C_2(\boldsymbol{u})$ is a copula for all $\lambda \in (0,1)$, i.e., convex combinations of copulas are copulas.

Proof (analytic).

1) Let $\boldsymbol{u}_j = (u_1, \dots, u_{j-1}, 0, u_{j+1}, \dots, u_d)$. Then

$$C(\boldsymbol{u}_j) = \lambda C_1(\boldsymbol{u}_j) + (1-\lambda)C_2(\boldsymbol{u}_j) = \lambda \cdot 0 + (1-\lambda) \cdot 0 = 0$$

since C_1, C_2 are grounded. Hence, C is grounded.

2) Let $\boldsymbol{u}_j = (1, \dots, 1, u_j, 1, \dots, 1)$. Then

$$C(\boldsymbol{u}_j) = \lambda C_1(\boldsymbol{u}_j) + (1-\lambda)C_2(\boldsymbol{u}_j) = \lambda u_j + (1-\lambda)u_j = u_j$$

since C_1, C_2 have $\mathrm{U}[0,1]$ margins. Hence, C has $\mathrm{U}[0,1]$ margins.

3) $\Delta_{(\boldsymbol{a},\boldsymbol{b}]}C = \lambda \Delta_{(\boldsymbol{a},\boldsymbol{b}]}C_1 + (1-\lambda)\Delta_{(\boldsymbol{a},\boldsymbol{b}]}C_2 \geq 0$, so C is d-increasing. $\square$

Proof (probabilistic). Let $\boldsymbol{U}_k \sim C_k,\ k \in \{1,2\}$ and let $X \sim \mathrm{B}(1,\lambda)$, independent of each other. Furthermore, let

$$\boldsymbol{U} = \begin{cases} \boldsymbol{U}_1, & \text{if } X = 1, \\ \boldsymbol{U}_2, & \text{if } X = 0. \end{cases}$$

The Law of Total Probability implies

$$\begin{aligned} \mathbb{P}(\boldsymbol{U} \le \boldsymbol{u}) &= \mathbb{P}(\boldsymbol{U} \le \boldsymbol{u}, X = 1) + \mathbb{P}(\boldsymbol{U} \le \boldsymbol{u}, X = 0) \\ &= \mathbb{P}(\boldsymbol{U}_1 \le \boldsymbol{u}, X = 1) + \mathbb{P}(\boldsymbol{U}_2 \le \boldsymbol{u}, X = 0) \\ &= \mathbb{P}(\boldsymbol{U}_1 \le \boldsymbol{u})\,\mathbb{P}(X = 1) + \mathbb{P}(\boldsymbol{U}_2 \le \boldsymbol{u})\,\mathbb{P}(X = 0) \\ &= C_1(\boldsymbol{u})\,\lambda + C_2(\boldsymbol{u})\,(1-\lambda) = C(\boldsymbol{u}). \end{aligned}$$

So $\boldsymbol{U} \sim C$ and hence C is a df. From the same calculation it follows that $\boldsymbol{U}$ has uniform margins, hence C is a copula. $\square$

Preliminaries

$T \nearrow$ means that T is *increasing*; $T \uparrow$ means that T is *strictly increasing*; $\operatorname{ran} T = \{T(x) : x \in \mathbb{R}\}$ denotes the *range of* T.

Proposition 7.3 (Working with generalized inverses)

Let $T : \mathbb{R} \to \mathbb{R} \nearrow$ with $T(-\infty) = \lim_{x \downarrow -\infty} T(x)$ and $T(\infty) = \lim_{x \uparrow \infty} T(x)$ and let $x, y \in \mathbb{R}$. Then,

(GI1) $T^-(y) = -\infty$ if and only if $T(x) \ge y$ for all $x \in \mathbb{R}$. Similarly, $T^-(y) = \infty$ if and only if $T(x) < y$ for all $x \in \mathbb{R}$.

(GI2) $T^- \nearrow$. If $T^-(y) \in (-\infty, \infty)$, T^- is left-continuous at y and admits a limit from the right at y.

(GI3) $T^-(T(x)) \le x$. If $T \uparrow$, then $T^-(T(x)) = x$.

(GI4) Let T be right-continuous. $T^-(y) < \infty$ implies $T(T^-(y)) \ge y$. Furthermore, $y \in \operatorname{ran} T \cup \{\inf T, \sup T\}$ implies $T(T^-(y)) = y$.

Moreover, if $y < \inf T$ then $T(T^-(y)) > y$ and if $y > \sup T$ then $T(T^-(y)) < y$.

(GI5) $T(x) \geq y$ implies $x \geq T^-(y)$. The other implication holds if T is right-continuous. Furthermore, $T(x) < y$ implies $x \leq T^-(y)$.

(GI6) $(T^-(y-), T^-(y+)) \subseteq \{x \in \mathbb{R} : T(x) = y\} \subseteq [T^-(y-), T^-(y+)]$, where $T^-(y-) = \lim_{z\uparrow y} T^-(z)$ and $T^-(y+) = \lim_{z\downarrow y} T^-(z)$.

(GI7) T is continuous if and only if $T^- \uparrow$ on $[\inf T, \sup T]$.
$T \uparrow$ if and only if T^- is continuous on $\operatorname{ran} T$.

(GI8) If T_1 and T_2 are right-continuous transformations with properties as T, then $(T_1 \circ T_2)^- = T_2^- \circ T_1^-$.

Proof. See Embrechts and Hofert (2013a). □

It is not important to know them, but rather to be aware of these "rules" when manipulating probabilities.

Lemma 7.4 (Probability transformation)
Let $X \sim F$, F continuous. Then $F(X) \sim \mathrm{U}[0,1]$.

Proof. Note that the *range of a rv X* is defined by

$$\operatorname{ran} X = \{x \in \mathbb{R} : \mathbb{P}(X \in (x-h, x]) > 0 \text{ for all } h > 0\}.$$

Since F is continuous on $\mathbb{R}$, (GI7) implies that $F^- \uparrow$ on $[\inf F, \sup F] = [0,1]$. Thus

$$\begin{aligned} \mathbb{P}(F(X) \le u) \underset{\text{(GI7)}}{=} \mathbb{P}(F^-(F(X)) \le F^-(u)) &\underset{\text{(GI3)}}{=} \mathbb{P}(X \le F^-(u)) \\ = F(F^-(u)) \underset{\text{(GI4)}}{=} u, \quad u \in [0,1], \end{aligned}$$

where (GI3) applies since $F \uparrow$ on $\operatorname{ran} X$. $\square$

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

Lemma 7.5 (Quantile transformation)

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

Proof. $\mathbb{P}(F^{-}(U) \le x) \underset{\text{(GI5)}}{=} \mathbb{P}(U \le F(x)) = F(x),\ x \in \mathbb{R}.$ □

- Probability and quantile transformations are the key to all applications involving copulas, e.g., goodness-of-fit testing (probability transformation) or simulation (quantile transformation). They will allow us to go from $\mathbb{R}^d$ to $[0,1]^d$ and back, respectively.
- Both transformations have multivariate equivalents (but then involving the underlying dependence structure as well!).

Sklar's Theorem

Theorem 7.6 (Sklar's Theorem)

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

$$H(x_1, \dots, x_d) = C(F_1(x_1), \dots, F_d(x_d)), \quad \boldsymbol{x} \in \mathbb{R}^d. \tag{35}$$

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

$$C(u_1, \dots, u_d) = H(F_1^-(u_1), \dots, F_d^-(u_d)), \quad \boldsymbol{u} \in \prod_{j=1}^d \operatorname{ran} F_j.$$

2) Conversely, given any copula C and univariate dfs $F_1, \dots, F_d$, H defined by (35) is a df with margins $F_1, \dots, F_d$.

Proof.

1) Proof for continuous $F_1, \dots, F_d$ only. Let $\boldsymbol{X} \sim H$ and define $U_j = F_j(X_j)$, $j \in \{1, \dots, d\}$. By the probability transformation, $U_j \sim \mathrm{U}[0, 1]$ (continuity!), $j \in \{1, \dots, d\}$, so the df C of $\boldsymbol{U}$ is a copula. Since $F_j \uparrow$ on $\operatorname{ran} X_j$, (GI3) implies that $X_j = F_j^-(F_j(X_j)) \overset{\text{a.s.}}{=} F_j^-(U_j)$, $j \in \{1, \dots, d\}$. Therefore,

$$
\begin{aligned}
H(\boldsymbol{x}) &= \mathbb{P}(X_j \le x_j \ \forall\, j) = \mathbb{P}(F_j^-(U_j) \le x_j \ \forall\, j) \underset{\text{(GI5)}}{=} \mathbb{P}(U_j \le F_j(x_j) \ \forall\, j) \\
&= C(F_1(x_1), \dots, F_d(x_d)), \quad \boldsymbol{x} \in \mathbb{R}^d.
\end{aligned}
$$

Hence C is a copula and satisfies (35).

(GI4) implies that $F_j(F_j^-(u_j)) = u_j$ for all $u_j \in \operatorname{ran} F_j$, so

$$
\begin{aligned}
C(u_1, \dots, u_d) &= C(F_1(F_1^-(u_1)), \dots, F_d(F_d^-(u_d))) \\
&\underset{(35)}{=} H(F_1^-(u_1), \dots, F_d^-(u_d)), \quad \boldsymbol{u} \in \prod_{j=1}^d \operatorname{ran} F_j.
\end{aligned}
$$

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

$$\mathbb{P}(\boldsymbol{X} \le \boldsymbol{x}) = \mathbb{P}(F_j^-(U_j) \le x_j\ \forall j) \underset{\text{(GI5)}}{=} \mathbb{P}(U_j \le F_j(x_j)\ \forall j)$$
$$= C(F_1(x_1), \dots, F_d(x_d)), \quad \boldsymbol{x} \in \mathbb{R}^d.$$

Therefore, H defined by (35) is a df (that of $\boldsymbol{X}$), with (by the quantile transformation) margins $F_1, \dots, F_d$. $\square$

Remark 7.7

- The general proof of Part 1) of Sklar's Theorem can be found, e.g., in Rüschendorf (2009). It works similarly but utilizes the *generalized distributional transform* $\boldsymbol{U} = (U_1, \dots, U_d) = (F_1(X_1, V_1), \dots, F_d(X_d, V_d))$, where $V_1, \dots, V_d \sim \mathrm{U}[0,1]$ are independent of $\boldsymbol{X}$ and
$$F_j(x, v) = \mathbb{P}(X < x) + v\mathbb{P}(X = x) = F(x-) + v(F(x) - F(x-)).$$
One can show: $\boldsymbol{U} \sim \mathrm{U}[0,1]^d$; $(F_1^-(U_1), \dots, F_d^-(U_d)) \overset{\text{a.s.}}{=} \boldsymbol{X}$.
- Non-uniqueness follows from different choices of $\boldsymbol{V}$, e.g., independent components or $\boldsymbol{V} = (V, \dots, V)$. See also the following example.

Example 7.8 (Bivariate Bernoulli distribution)

Let (X_1, X_2) follow a bivariate Bernoulli distribution with $\mathbb{P}(X_1 = k, X_2 = l) = 1/4,\ k, l \in \{0, 1\}$. $\Rightarrow \mathbb{P}(X_j = k) = 1/2,\ k \in \{0, 1\}$, $\operatorname{ran} F_j = \{0, 1/2, 1\},\ j \in \{1, 2\}$. Any copula with $C(1/2, 1/2) = 1/4$ satisfies (35) (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 $\delta(u) = u^2$). See Genest and Nešlehová (2007) for more details on copulas for discrete data.

Interpretation of Sklar's Theorem

1) Allows one to decompose any df H into its margins and a copula. This allows one to study multivariate distributions independently of the margins (interesting for statistical applications, e.g., parameter estimation or goodness-of-fit).

2) Allows one to compose new multivariate distributions (interesting for constructing flexible models, sampling, stress testing).

- A *copula model* for $\boldsymbol{X}$ means $H(\boldsymbol{x}) = C(F_1(x_1), \dots, F_d(x_d))$ for some (parametric) copula C and (parametric) marginals $F_1, \dots, F_d$.
- $\boldsymbol{X}$ (or H) with margins $F_1, \dots, F_d$ *has copula* C if (35) holds.

Invariance principle

Lemma 7.9 (Core of the invariance principle)
Let $X_j \sim F_j$, F_j continuous, $j \in \{1, \dots, d\}$. Then

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

Proof.

$$\text{“}\Rightarrow\text{”} \;\; \mathbb{P}(F_j(X_j) \le u_j \,\forall\, j) \underset{\text{cont.}}{=} \mathbb{P}(F_j(X_j) < u_j \,\forall\, j) \underset{\text{(GI5)}}{=} \mathbb{P}(X_j < F_j^-(u_j) \,\forall\, j)$$
$$\underset{\text{cont.}}{=} \mathbb{P}(X_j \le F_j^-(u_j) \,\forall\, j) = H(F_1^-(u_1), \dots, F_d^-(u_d)) \underset{\text{Sklar}}{=} C(\boldsymbol{u}).$$

"$\Leftarrow$" Since $F_j \uparrow$ on $\operatorname{ran} X_j$, $j \in \{1, \dots, d\}$,

$$H(\boldsymbol{x}) \underset{\text{(GI3)}}{=} \mathbb{P}(F_j^-(F_j(X_j)) \le x_j \,\forall\, j) \underset{\text{(GI5)}}{=} \mathbb{P}(F_j(X_j) \le F_j(x_j) \,\forall\, j)$$
$$\underset{\text{ass.}}{=} C(F_1(x_1), \dots, F_d(x_d)) \quad \underset{\text{Sklar}}{\Rightarrow} \quad \boldsymbol{X} \text{ has copula } C \qquad \square$$

Theorem 7.10 (Invariance principle)
Let $\boldsymbol{X} \sim H$ with continuous margins $F_1, \dots, F_d$ and copula C. If $T_j \uparrow$ on $\operatorname{ran} X_j$ 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 $\operatorname{ran} X_j$ and X_j is continuously distributed, $T_j(X_j)$ is continuously distributed and we have

$$F_{T_j(X_j)}(x) = \mathbb{P}(T_j(X_j) \le x) = \mathbb{P}(T_j(X_j) < x) \underset{\text{(GI5)}}{=} \mathbb{P}(X_j < T_j^-(x))$$
$$= \mathbb{P}(X_j \le T_j^-(x)) = F_j(T_j^-(x)), \quad x \in \mathbb{R}.$$

This implies that

$$\begin{aligned}\mathbb{P}(F_{T_j(X_j)}(T_j(X_j)) \le u_j\,\forall\, j) &= \mathbb{P}(F_j(T_j^-(T_j(X_j))) \le u_j\,\forall\, j)\\ &\underset{\text{(GI3)}}{=} \mathbb{P}(F_j(X_j) \le u_j\,\forall\, j) \underset{\text{“only if”}}{\overset{\text{L.7.9}}{=}} C(\boldsymbol{u}).\end{aligned}$$

The claim follows by the if part ("$\Leftarrow$") of Lemma 7.9. $\square$

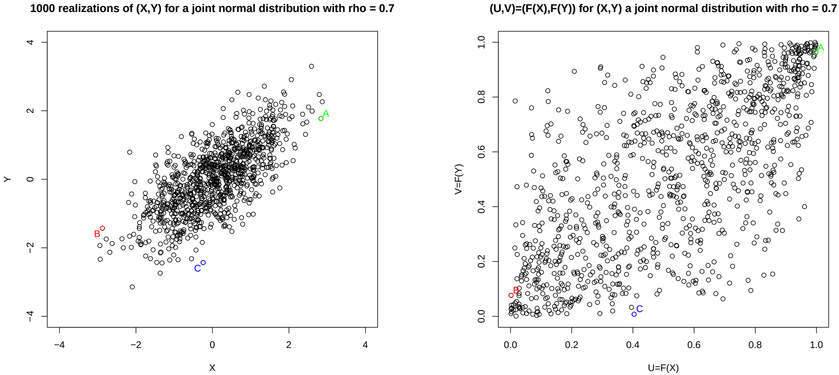
- The invariance principle allows us to study dependence in terms of the margin-free $\boldsymbol{U} = (F_1(X_1), \dots, F_d(X_d))$ instead of $\boldsymbol{X} = (X_1, \dots, X_d)$, $\boldsymbol{X}$, $\boldsymbol{U}$ have the same copula!

How does $(F_1(\cdot), \dots, F_d(\cdot))$ act on $\boldsymbol{X}$? A picture is worth a thousand words...

Visualizing the first part of Sklar's Theorem

Left: Scatter plot of $n = 1000$ samples from $(X_1, X_2) \sim \mathrm{N}_2(\mathbf{0}, P)$, where $P = \begin{pmatrix} 1 & 0.7 \\ 0.7 & 1 \end{pmatrix}$. We mark three points A, B, C.

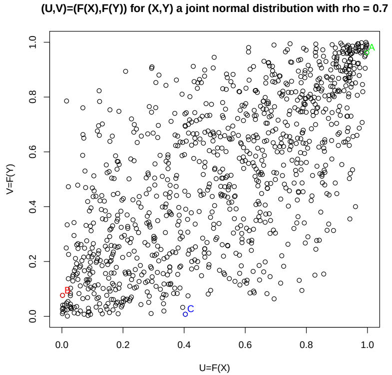
Right: After applying the F_j'sy (the df Φ of $\mathrm{N}(0,1)$), scatter plot of the corresponding Gauss copula. Note how the points A, B, C change.

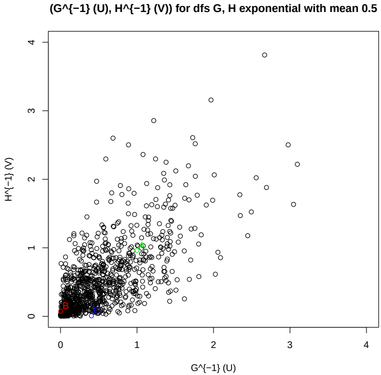


Visualizing the second part of Sklar's Theorem

Left: Same Gauss copula scatter plot as before. Apply marginal $\mathrm{Exp}(2)$-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.

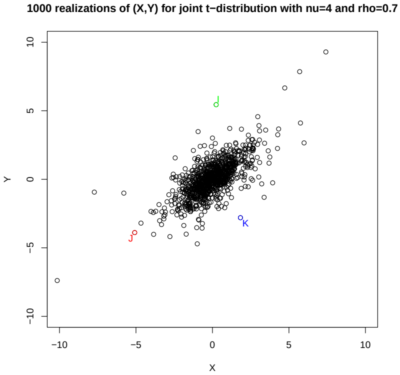


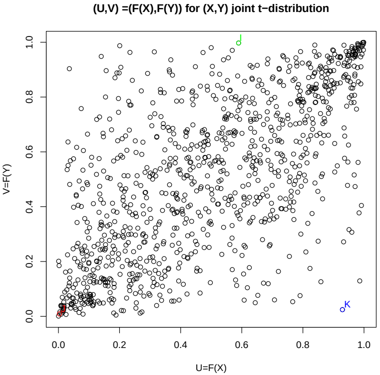


Visualizing the first part of Sklar's Theorem

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: After applying the F_j's (the df of t_4), scatter plot of the corresponding t_4 copula. Note how the points I, J, K change.

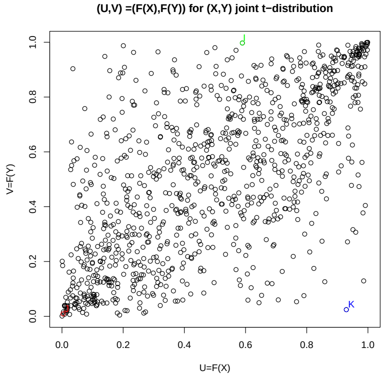


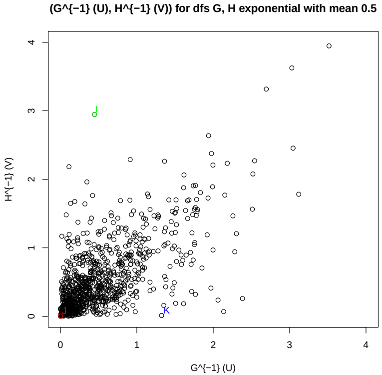


Visualizing the second part of Sklar's Theorem

Left: Same t_4 copula scatter plot as before. Apply marginal $\mathrm{Exp}(2)$-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 7.11 (Fréchet–Hoeffding bounds)

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

1) For any d-dimensional copula C,

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

2) W is a copula if and only if $d = 2$.
3) M is a copula for all $d \ge 2$.

Proof.

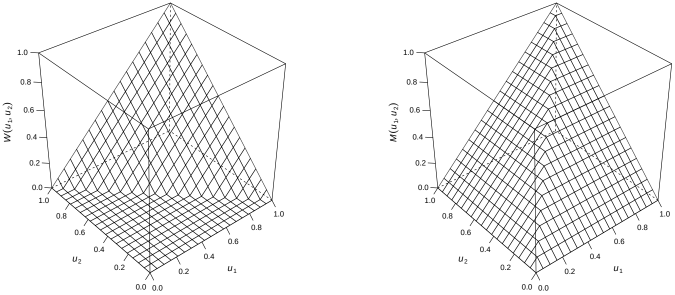
1)
- By (34), $1 - C(\boldsymbol{u}) = C(\boldsymbol{1}) - C(\boldsymbol{u}) \le \sum_{j=1}^{d}(1 - u_j) = d - \sum_{j=1}^{d} u_j$, so $C(\boldsymbol{u}) \ge \sum_{j=1}^{d} u_j - d + 1$. Also, $C(\boldsymbol{u}) \ge 0$. So $C(\boldsymbol{u}) \ge W(\boldsymbol{u})$.
- Since copulas are componentwise increasing, $C(\boldsymbol{u}) \le C(1,\dots,1,u_j,1,\dots,1) = u_j$ for all j. Hence, $C(\boldsymbol{u}) \le \min_{1\le j\le d}\{u_j\} = M(\boldsymbol{u})$.

2) W is a copula for $d=2$ since $(U,1-U)\sim W$ for $U\sim \mathrm{U}[0,1]$. W is not a copula for $d\ge 3$ since

$$
\begin{aligned}
&\Delta_{(\frac{1}{2},\mathbf{1}]}W\\
&=\sum_{\boldsymbol{i}\in\{0,1\}^d}(-1)^{\sum_{j=1}^d i_j}W\big(\tfrac{1}{2}^{i_1},\dots,\tfrac{1}{2}^{i_d}\big)\\
&=\max\{1+1+1+\dots+1-d+1,0\}\quad (i_j=0\ \forall\, j)\\
&\quad -d\max\{\tfrac{1}{2}+1+1+\dots+1-d+1,0\}\quad (\exists!\, j: i_j=1)\\
&\quad +\binom{d}{2}\max\{\tfrac{1}{2}+\tfrac{1}{2}+1+\dots+1-d+1,0\}\ \ (\exists!\ \text{two } j: i_j=1)\\
&\quad -\dots+(-1)^d\max\{\tfrac{1}{2}+\dots+\tfrac{1}{2}-d+1,0\}\quad (i_j=1\ \forall\, j)\\
&=1-\tfrac{d}{2}<0\quad \text{for } d\ge 3.
\end{aligned}
$$

3) M is a copula for all $d\ge 2$ since $(U,\dots,U)\sim M$ for $U\sim\mathrm{U}[0,1]$. $\square$

- 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 7.18 later.
- The Fréchet–Hoeffding bounds lead to bounds for any df H, via

$$\max\Big\{\sum_{j=1}^{d} F_j(x_j) - d + 1, 0\Big\} \le H(\boldsymbol{x}) \le \min_{1 \le j \le d}\{F_j(x_j)\}.$$

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 H via Sklar's Theorem;
- *Explicit copulas*: have simple closed-from expressions and follow construction principles of copulas.

Fundamental copulas

- $\Pi(\boldsymbol{u}) = \prod_{j=1}^{d} u_j$ is the *independence copula* since $C(F_1(x_1), \dots, F_d(x_d)) \underset{\text{Sklar}}{=} H(\boldsymbol{x}) \underset{\text{ind.}}{=} \prod_{j=1}^{d} F_j(x_j)$ if and only if $C(\boldsymbol{u}) = \Pi(\boldsymbol{u})$ (now replace x_j by $F_j^{-}(u_j)$ 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.) $\boldsymbol{X} \sim \mathrm{N}_d(\mathbf{0}, P)$. The *Gauss copula* (family) is given by

$$C_P^{\mathrm{Ga}}(\boldsymbol{u}) = \mathbb{P}(\Phi(X_1) \le u_1, \dots, \Phi(X_d) \le u_d)$$
$$= \Phi_P(\Phi^{-1}(u_1), \dots, \Phi^{-1}(u_d))$$

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

- $P = I_d \Rightarrow C = \Pi$; and $P = J_d = \mathbf{1}\mathbf{1}^\top \Rightarrow C = M$;

  $d = 2$ and $\rho = P_{12} = -1 \Rightarrow C = W$.
- For $d > 3$, C_P^{Ga} is evaluated by (randomized quasi-)Monte Carlo.
- Sklar's Theorem $\Rightarrow$ The density of $C(\boldsymbol{u}) = H(F_1^-(u_1), \dots, F_d^-(u_d))$ is

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

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

$$c_P^{\mathsf{Ga}}(\boldsymbol{u}) = \frac{1}{\sqrt{\det P}} \exp\Big(-\frac{1}{2}\boldsymbol{x}^\top (P^{-1} - I_d)\boldsymbol{x}\Big), \tag{36}$$

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

t copulas ("lonely island copula")

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

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

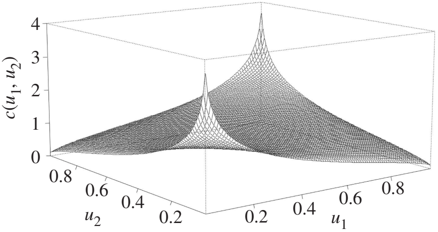
where $t_{\nu,P}$ is the df of $t_d(\nu,\mathbf{0},P)$ and t_ν the df of the univariate t distribution with ν degrees of freedom.

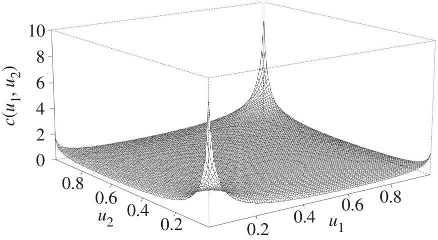
- $P = J_d = \mathbf{1}\mathbf{1}^\top \Rightarrow C = M$; and $d = 2$ and $\rho = P_{12} = -1 \Rightarrow C = W$. However, $P = I_d \Rightarrow C \neq \Pi$ (unless $\nu = \infty$ in which case $C^t_{\nu,P} = C^{\mathrm{Ga}}_P$).
- For $d > 3$, $C^t_{\nu,P}$ is evaluated by (randomized quasi-)Monte Carlo.
- Sklar's Theorem $\Rightarrow$ The density of $C^t_{\nu,P}$ is

$$c^t_{\nu,P}(\boldsymbol{u}) = \frac{\Gamma((\nu+d)/2)}{\Gamma(\nu/2)\sqrt{\det P}}\left(\frac{\Gamma(\nu/2)}{\Gamma((\nu+1)/2)}\right)^d \frac{(1+\boldsymbol{x}^\top P^{-1}\boldsymbol{x}/\nu)^{-(\nu+d)/2}}{\prod_{j=1}^d (1+x_j^2/\nu)^{-(\nu+1)/2}},$$

  for $\boldsymbol{x} = (t_\nu^{-1}(u_1),\dots,t_\nu^{-1}(u_d))$.
- 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 $C^{\mathrm{Ga}}_{\rho=0.3}$ (left) and $C^{t}_{4,\,\rho=0.3}(\boldsymbol{u})$ (right).





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

Advantages:

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

Drawbacks:

- Typically, C is not explicit
- Radially symmetric, so same lower/upper tail behavior ($\lambda_L = \lambda_U$)

Explicit copulas

Archimedean copulas are copulas of the form

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

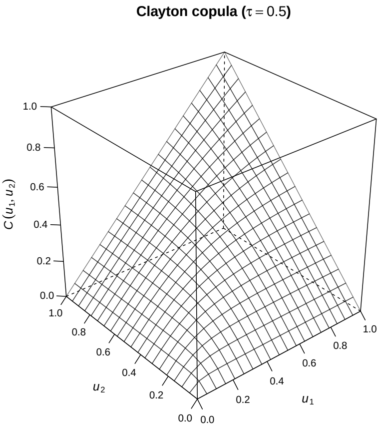
where the *(Archimedean) generator* $\psi : [0,\infty) \to [0,1]$ is $\downarrow$ on $[0, \inf\{t : \psi(t) = 0\}]$ and satisfies $\psi(0) = 1$, $\psi(\infty) = \lim_{t\to\infty} \psi(t) = 0$; we set $\psi^{-1}(0) = \inf\{t : \psi(t) = 0\}$. The set of all generators is denoted by Ψ. If $\psi(t) > 0$,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

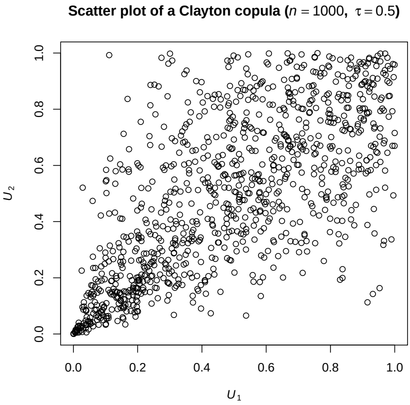
Examples

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

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

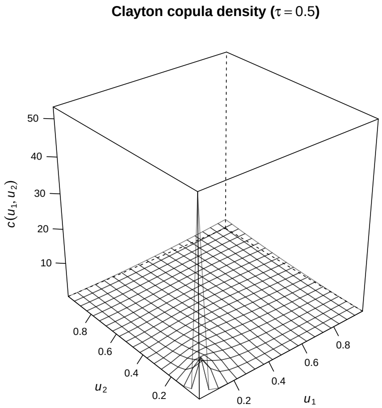
Right: Corresponding scatter plot (sample size $n = 1000$)

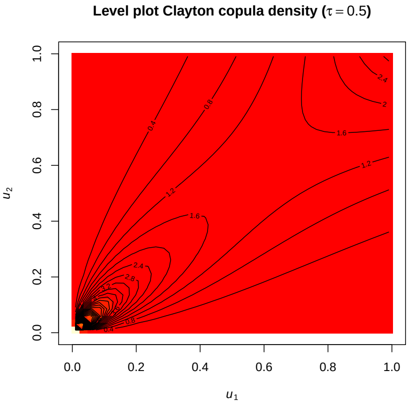




Left: Plot of the corresponding density.

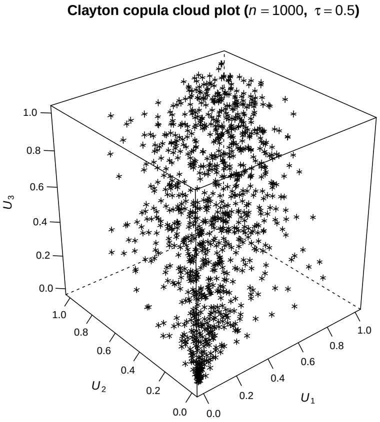
Right: Level plot of the density (with heat colors).

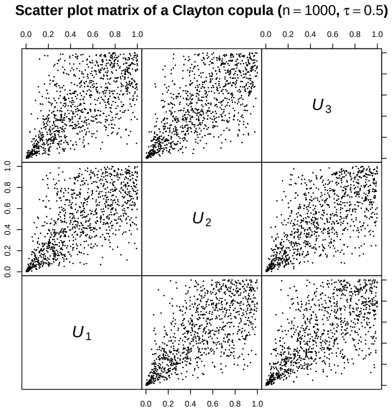




Left: Cloud plot of a trivariate Clayton copula (sample size $n = 1000$; Kendall's tau 0.5).

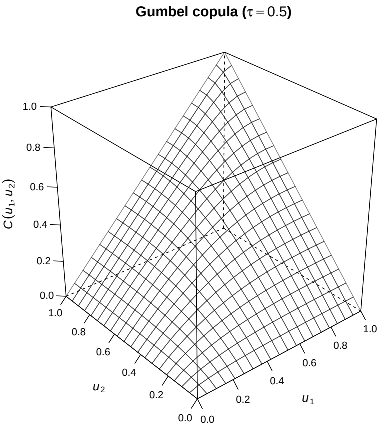
Right: Corresponding scatter plot matrix.

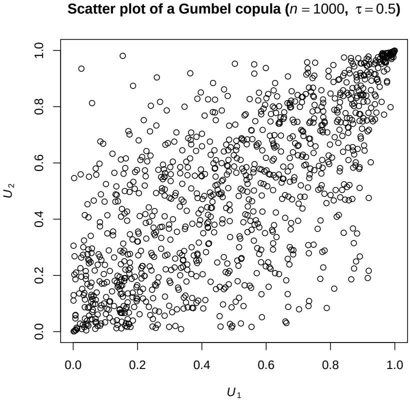




Left: Plot of a bivariate Gumbel copula (Kendall's tau 0.5).

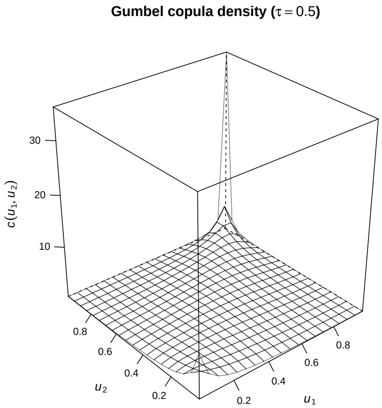
Right: Corresponding scatter plot (sample size $n = 1000$)

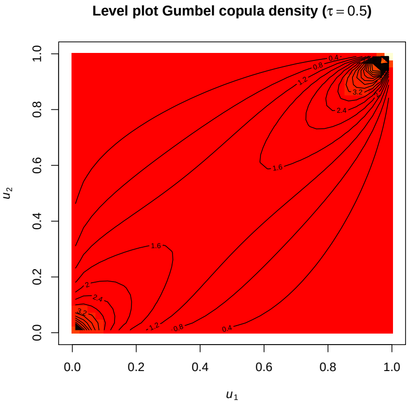




Left: Plot of the corresponding density.

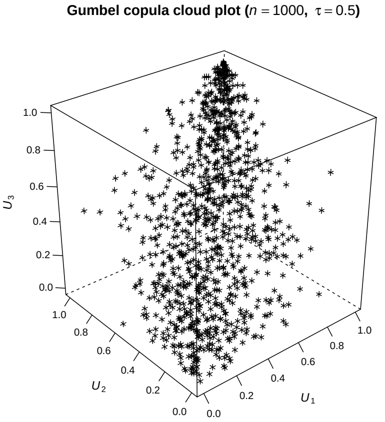
Right: Level plot of the density (with heat colors).

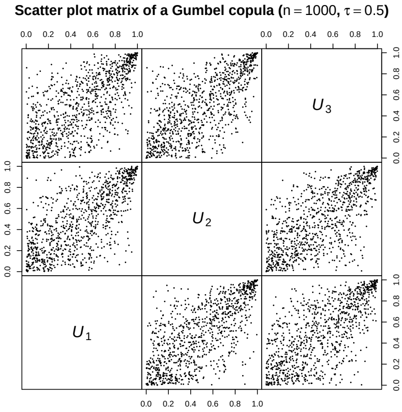




Left: Cloud plot of a trivariate Gumbel copula (sample size $n = 1000$; Kendall's tau 0.5).

Right: Corresponding scatter plot matrix.





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 ($\lambda_L \neq \lambda_U$ possible)

Drawbacks:

- Exchangeable, i.e., (functionally) symmetric (all pairs have the same dependence)
- Often used only with a small number of parameters (some extensions available, but still less than $d(d-1)/2$)

Excursion: Other animals in the zoo (of copulas)

Extreme value copulas

- *Extreme value copulas* are the copulas C of limiting distributions of properly location-scale transformed componentwise maxima of a sequence of random vectors.
- They are given by

$$C(\boldsymbol{u}) = \Bigl(\prod_{j=1}^{d} u_j\Bigr)^{A\left(\frac{\log u_1}{\log \Pi(\boldsymbol{u})},\dots,\frac{\log u_d}{\log \Pi(\boldsymbol{u})}\right)}$$

  for a *Pickands dependence function* A; see Ressel (2013) for a characterization of A.
- Examples: Gumbel copula, Marshall-Olkin copulas.
- For more details, see Jaworski et al. (2010, Chapter 6).

Marshall–Olkin copulas

- Here we focus on $d = 2$ only.
- **Motivation:** Consider a system of two components affected by three types of fatal shocks: One hitting the first, one the second, and one both components simultaneously. Assuming the shocks to follow independent homogeneous Poisson processes, the times of occurrence are given by $T_j \sim \mathrm{Exp}(\lambda_j)$, $j \in \{1, 2\}$, $T_{12} \sim \mathrm{Exp}(\lambda_{12})$ (all independent). The lifetimes of the components are thus

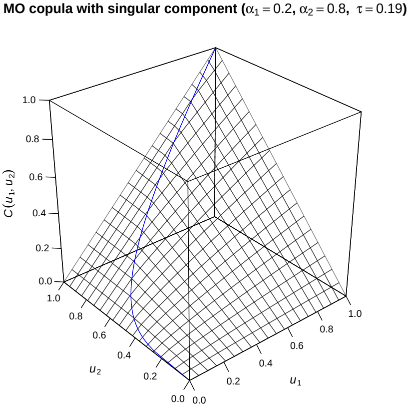
$$X_j = \min\{T_j, T_{12}\}, \quad j \in \{1, 2\}.$$

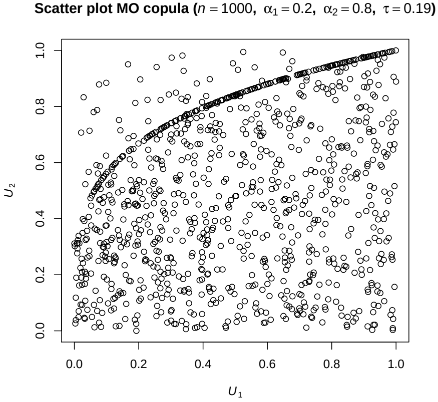
- The survival copula $\hat{C}$ (see later) of $\boldsymbol{X}$ is known as *Marshall–Olkin copula*. Derivation:

$$\begin{aligned}\bar{H}(x_1, x_2) &= \mathbb{P}(X_1 > x_1, X_2 > x_2) \\ &= \mathbb{P}(T_1 > x_1, T_2 > x_2, T_{12} > \max\{x_1, x_2\}) \\ &= \exp(-\lambda_1 x_1 - \lambda_2 x_2 - \lambda_{12} \max\{x_1, x_2\}),\end{aligned}$$

from which we obtain that $\bar{F}_j(x) = \exp(-(\lambda_j + \lambda_{12})x)$, $j \in \{1, 2\}$. With $\alpha_j = \frac{\lambda_{12}}{\lambda_j + \lambda_{12}}$, $j \in \{1, 2\}$, it follows that

$$\hat{C}(u_1, u_2) \underset{\text{Sklar}}{=} \bar{H}(\bar{F}_1^-(u_1), \bar{F}_2^-(u_2)) = u_1^{1-\alpha_1} u_2^{1-\alpha_2} \min\{u_1^{\alpha_1}, u_2^{\alpha_2}\}$$
$$= \min\{u_1 u_2^{1-\alpha_2}, u_1^{1-\alpha_1} u_2\}, \quad \alpha_1, \alpha_2 \in [0, 1].$$





- Any copula C can be decomposed into

$$C(\boldsymbol{u}) = A_C(\boldsymbol{u}) + S_C(\boldsymbol{u}) \quad \text{(Lebesgue Decomposition)},$$

where

$$A_C(\boldsymbol{u}) = \int_{(\boldsymbol{0},\boldsymbol{u}]} D_{d,\dots,1}C(\boldsymbol{v})\,d\boldsymbol{v}$$

is the *absolutely continuous component* of C and

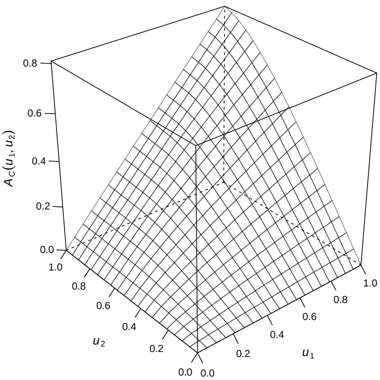
$$S_C(\boldsymbol{u}) = C(\boldsymbol{u}) - A_C(\boldsymbol{u})$$

is the *singular component* of C.

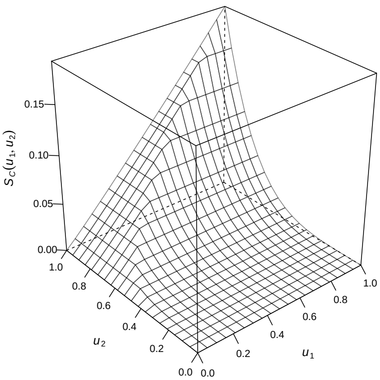
- For Marshall–Olkin copulas, integrating $D_{21}C(u_1, u_2)$ (exists for $u_1^{\alpha_1} \neq u_2^{\alpha_2}$) yields A_C and thus S_C. The mass on the singular component is $\mathbb{P}(U_1^{\alpha_1} = U_2^{\alpha_2}) = S_C(1,1) = \frac{\alpha_1\alpha_2}{\alpha_1+\alpha_2-\alpha_1\alpha_2}$.

Absolutely continuous and singular components of a Marshall–Olkin copula:

Abs. cont. comp. A_C **of a MO copula** ($\alpha_1 = 0.2$, $\alpha_2 = 0.8$, $\tau = 0.19$)

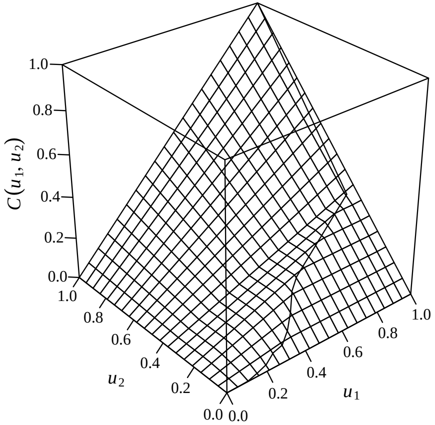


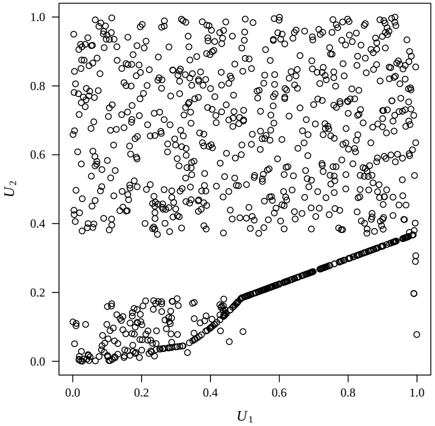
Singular comp. S_C **of a MO copula** ($\alpha_1 = 0.2$, $\alpha_2 = 0.8$, $\tau = 0.19$)



A rather exotic example (with singular component)

A Sibuya copula; see Hofert and Vrins (2013).





7.1.3 Meta distributions

- *Fréchet class*: Class of all dfs H with given marginal dfs $F_1, \dots, F_d$; *Meta-C models*: All dfs H with the same given copula C.
- **Example:** A meta-Gauss model is a multivariate df H with Gauss copula C and some margins $F_1, \dots, F_d$. Such a model, with exponential margins, is used in Li's model in credit risk.

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 7.9.

Algorithm 7.12 (Simulation of implicit copulas)

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

Example 7.13

- Sampling Gauss copulas C_P^{Ga}:
  1) Sample $\boldsymbol{X} \sim \mathrm{N}_d(\mathbf{0}, P)$ ($\boldsymbol{X} \overset{\mathrm{d}}{=} A\boldsymbol{Z}$ for $AA^\top = P$, $\boldsymbol{Z} \sim \mathrm{N}_d(\mathbf{0}, I_d)$).
  2) Return $\boldsymbol{U} = (\Phi(X_1), \dots, \Phi(X_d))$.
- Sampling t_ν copulas $C_{\nu,P}^t$:
  1) Sample $\boldsymbol{X} \sim t_d(\nu, \mathbf{0}, P)$ ($\boldsymbol{X} \overset{\mathrm{d}}{=} \sqrt{W} A\boldsymbol{Z}$ for $W = \frac{1}{V}$, $V \sim \Gamma(\frac{\nu}{2}, \frac{\nu}{2})$).
  2) Return $\boldsymbol{U} = (t_\nu(X_1), \dots, t_\nu(X_d))$.

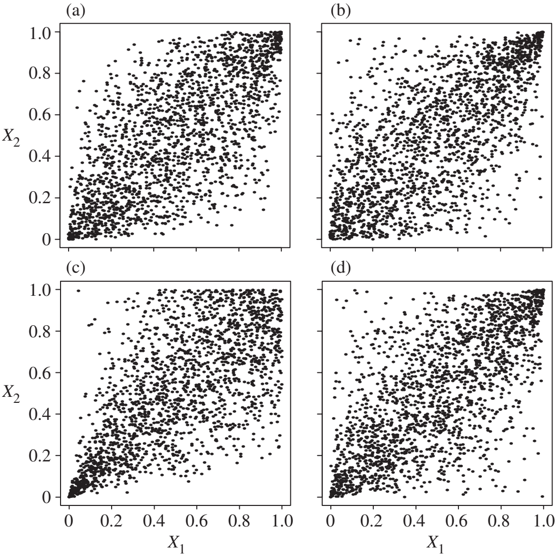
Sampling meta distributions

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

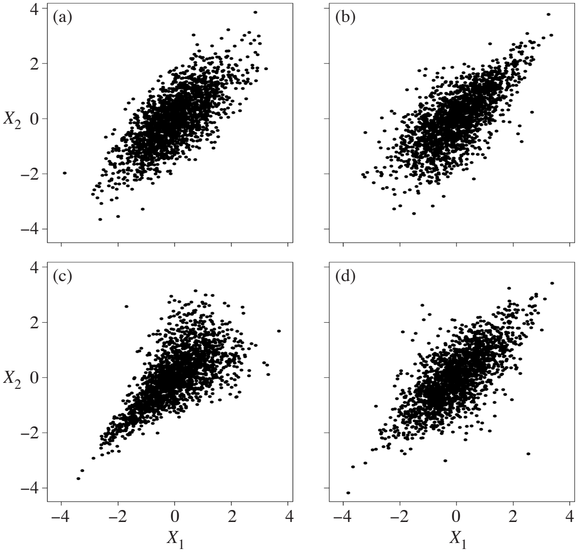
Algorithm 7.14 (Sampling)

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

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



. . . transformed to $\mathrm{N}(0,1)$ 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 7.15 (Conditional distribution method)
If C is a d-dimensional copula and $\boldsymbol{U}' \sim \mathrm{U}[0,1]^d$, let

$$\begin{aligned} U_1 &= U_1', \\ U_2 &= C^-(U_2' \,|\, U_1), \\ &\ \ \vdots \\ U_d &= C^-(U_d' \,|\, U_1, \dots, U_{d-1}). \end{aligned}$$

Then $\boldsymbol{U} \sim C$.

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

Theorem 7.16 (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

$$C(u_j \,|\, u_1, \ldots, u_{j-1}) = \frac{\mathrm{D}_{j-1,\ldots,1}\, C^{(1,\ldots,j)}(u_1, \ldots, u_j)}{\mathrm{D}_{j-1,\ldots,1}\, C^{(1,\ldots,j-1)}(u_1, \ldots, u_{j-1})}$$

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

Example 7.17 (Conditional distribution method for Clayton copula)

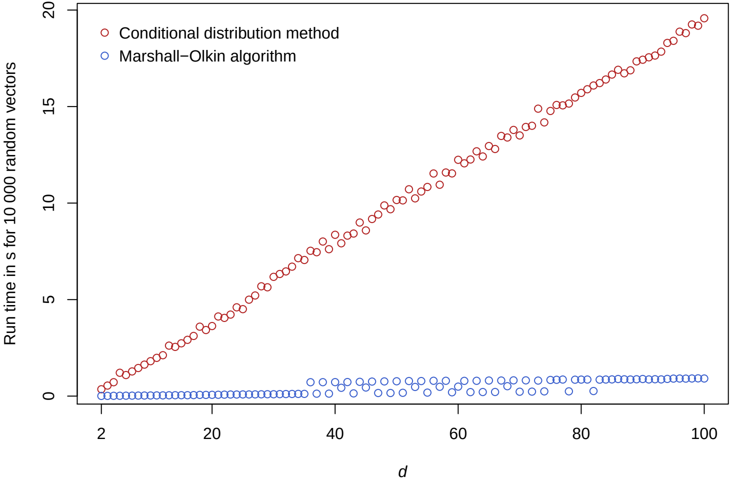
- For a Clayton copula, $\psi^{(k)}(t) = (-1)^k(1+t)^{-k+1/\theta}\prod_{l=0}^{k-1}(l+1/\theta)$.
- By Theorem 7.16, $C(u_j \,|\, u_1, \ldots, u_{j-1}) = \frac{\psi^{(j-1)}\left(\sum_{k=1}^{j}\psi^{-1}(u_k)\right)}{\psi^{(j-1)}\left(\sum_{k=1}^{j-1}\psi^{-1}(u_k)\right)}$
$= \left(\frac{1-(j-1)+\sum_{k=1}^{j-1}u_k^{-\theta}}{1-j+\sum_{k=1}^{j}u_k^{-\theta}}\right)^{j-1+1/\theta}.$

- Thus $C^{-}(u_j \,|\, u_1, \ldots, u_{j-1})$ equals

$$\left(1 + \left(1 - (j-1) + \sum_{k=1}^{j-1} u_k^{-\theta}\right)\left(u_j^{-1/(j-1+1/\theta)} - 1\right)\right)^{-1/\theta}$$

- The Clayton copula is one of the rare cases where both $C(u_j \,|\, u_1, \ldots, u_{j-1})$ and $C^{-}(u_j \,|\, u_1, \ldots, u_{j-1})$ can be computed explicitly. These quantities are often difficult to compute for $d > 2$.
- For all well-known copula families, the conditional distribution method is neither simple to apply nor fast $\Rightarrow$ Efficient sampling algorithms are typically family-specific.

A comparison with the conditional distribution method (here: Clayton)



This is the best-case scenario for applying the conditional distribution method! But even here there are faster algorithms (see below).

7.1.5 Further properties of copulas

Survival copulas

- If $\boldsymbol{U} \sim C$, then $\mathbf{1} - \boldsymbol{U} \sim \hat{C}$, the *survival copula* of C.
- Survival copulas transform a copula into another copula (the one of $\mathbf{1} - \boldsymbol{U}$).
- $\hat{C}$ can be expressed as

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

  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, $\hat{c}(\boldsymbol{u}) = c(\mathbf{1} - \boldsymbol{u})$.

- 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 $\boldsymbol{X}$ is radially symmetric about $\boldsymbol{a}$ if and only if $C = \hat{C}$.
- Sklar's Theorem can also be formulated for survival functions. In this case, the main part reads

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

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

  $\Rightarrow$ Survival copulas combine marginal survival functions to joint survival functions. Note that $\hat{C}$ is a df, whereas $\bar{H}$ and $\bar{F}_1, \dots, \bar{F}_d$ are not!
- From this we derived Marshall–Olkin copulas.

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 h of H satisfies

$$h(\boldsymbol{x}) = c(F_1(x_1), \dots, F_d(x_d)) \prod_{j=1}^{d} f_j(x_j) \tag{37}$$

  As seen before, we can recover c via

$$c(\boldsymbol{u}) = \frac{h(F_1^{-1}(u_1), \dots, F_d^{-1}(u_d))}{f_1(F_1^{-1}(u_1)) \cdot \dots \cdot f_d(F_d^{-1}(u_d))}.$$

- It follows from (37) that the log-density splits into

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

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

Exchangeability

- $\boldsymbol{X}$ is *exchangeable* if

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

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

- A copula C is *exchangeable* if it is the df of an exchangeable $\boldsymbol{U}$ with $\mathrm{U}[0,1]$ margins. This holds if only if $C(u_1, \ldots, u_d) = C(u_{\pi(1)}, \ldots, u_{\pi(d)})$ for all possible permutations of arguments, i.e., if C is symmetric.

- Exchangeable/symmetric copulas are useful for modeling homogeneous portfolios or as (possibly crude) approximations of the underlying (possibly non-homogeneous) dependence structure.

- **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 7.18 (Perfect dependence)

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

Proof. We only consider Part 1) as Part 2) works similarly.

"$\Rightarrow$" By assumption, $\mathbb{P}(X_2 \le x)$ equals $\mathbb{P}(F_2^-(1-F_1(X_1)) \le x) \underset{\text{(GI5)}}{=} \mathbb{P}(1-F_1(X_1) \le F_2(x)) = F_2(x)$. If (X_1, X_2) has copula C, then

$$
\begin{aligned}
C(\boldsymbol{u}) &\underset{\text{“only if”}}{\overset{\text{L.7.9}}{=}} \mathbb{P}(F_1(X_1) \le u_1, F_2(F_2^-(1-F_1(X_1))) \le u_2) \\
&\underset{\text{(GI4)}}{=} \mathbb{P}(F_1(X_1) \le u_1, 1-F_1(X_1) \le u_2) \\
&= \mathbb{P}(1-u_2 < U \le u_1) = W(u_1, u_2) \quad \text{for } U \sim \mathrm{U}[0,1].
\end{aligned}
$$

"$\Leftarrow$" Let $\boldsymbol{u} = (u_1, u_2)$ and note that $W(\boldsymbol{u}) = 0$ for all $\boldsymbol{u} \in [0,1]^2$ such that $u_1 + u_2 - 1 < 0$, so W puts no mass below the secondary diagonal. Similarly (exercise!) one shows that W puts no mass above the diagonal. This implies that W puts mass 1 on the secondary diagonal. Since $F_2 \uparrow \operatorname{ran} X_2$, we thus obtain $\mathbb{P}(X_2 = F_2^-(1-F_1(X_1))) = \mathbb{P}(F_2(X_2) = F_2(F_2^-(1-F_1(X_1)))) \underset{\text{(GI4)}}{=} \mathbb{P}(F_2(X_2) = 1-F_1(X_1)) = \mathbb{P}(U_2 = 1-U_1) = 1$. $\square$

Comonotone additivity of quantiles

Proposition 7.19 (Comonotone additivity)
Let $\alpha \in (0,1)$ and $X_j \sim F_j$, $j \in \{1,\dots,d\}$, be comontone. Then $F^-_{X_1+\dots+X_d}(\alpha) = F_1^-(\alpha) + \dots + F_d^-(\alpha)$.

Proof. Consider $T(u) = F_1^-(u) + \dots + F_d^-(u) \nearrow$, left-continuous and let $U \sim \mathrm{U}[0,1]$. We first show that $F^-_{T(U)}(u) = T(u)$, for all $u \in [0,1]$.

1) T left-continuous $\Rightarrow T(u) \le x \Leftrightarrow \; u \le u_x := \sup\{u : T(u) \le x\}$

2) 1) $\Rightarrow \{T(U) \le x\} = \{U \le u_x\} \;\Rightarrow F_{T(U)}(x) = F_U(u_x) = u_x$.

$\Rightarrow F^-_{T(U)}(u) \le x \underset{\text{(GI5)}}{\Leftrightarrow} F_{T(U)}(x) \ge u \underset{2)}{\Leftrightarrow} u_x \ge u \underset{1)}{\Leftrightarrow} T(u) \le x$

$\Rightarrow$ Choosing $x = T(u)$ and $x = F^-_{T(U)}(u)$, we see that $F^-_{T(U)}(u) = T(u)$.
Now Proposition 7.18 1) implies that $(X_1,\dots,X_d) \overset{\mathrm{d}}{=} (F_1^-(U),\dots,F_d^-(U))$, so that $F^-_{\sum_{j=1}^d X_j}(\alpha) = F^-_{T(U)}(\alpha) = T(\alpha) = \sum_{j=1}^d F_j^-(\alpha)$. $\square$

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

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

Proposition 7.20 (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 H. Then

$$\mathrm{Cov}[X_1, X_2] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (H(x_1, x_2) - F_1(x_1)F_2(x_2))\, dx_1 dx_2.$$

Proof. Let (X_1', X_2') be an iid-copy of (X_1, X_2). Consider

$$\begin{aligned} &2\,\mathrm{Cov}[X_1, X_2] \\ &= \mathbb{E}[(X_1 - \mathbb{E}X_1)(X_2 - \mathbb{E}X_2)] + \mathbb{E}[(X_1' - \mathbb{E}X_1')(X_2' - \mathbb{E}X_2')] \end{aligned}$$

$$\underset{\text{check}}{=} \mathbb{E}[((X_1 - \mathbb{E}X_1) - (X_1' - \mathbb{E}X_1')) \cdot ((X_2 - \mathbb{E}X_2) - (X_2' - \mathbb{E}X_2'))]$$
$$= \mathbb{E}[(X_1 - X_1')(X_2 - X_2')].$$

With $b - a = \int_{-\infty}^{\infty}(\mathbb{1}_{\{a \le x\}} - \mathbb{1}_{\{b \le x\}})\,dx$ for all $a, b \in \mathbb{R}$, we obtain that

$$2\,\mathrm{Cov}[X_1, X_2]$$
$$= \mathbb{E}\Big[\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(\mathbb{1}_{\{X_1' \le x_1\}} - \mathbb{1}_{\{X_1 \le x_1\}})(\mathbb{1}_{\{X_2' \le x_2\}} - \mathbb{1}_{\{X_2 \le x_2\}})\,dx_1 dx_2\Big]$$
$$\underset{\text{Fubini}}{=} \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \mathbb{E}[\dots]\,dx_1 dx_2 \underset{\text{ind.}}{\overset{\text{multiply}}{=}} 2\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(H(x_1, x_2) - F_1(x_1)F_2(x_2))\,dx_1 dx_2.$$

$\square$

We now collect well-known and not so well-known properties of ρ (which also reveil ρ's deficiencies to quantify dependence).

Proposition 7.21 (Properties of linear correlation)

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

1) $|\rho| \leq 1$. Furthermore, $|\rho| = 1$ if and only if there are constants $a \in \mathbb{R}\backslash\{0\}, b \in \mathbb{R}$ with $X_2 = aX_1 + b$ a.s. with $a \gtrless 0$ if and only if $\rho = \pm 1$.
2) If X_1 and X_2 are independent, then $\rho = 0$ (the converse is not true in general, see Example 7.23 below).
3) ρ is invariant under strictly increasing linear transformations on $\operatorname{ran} X_1 \times \operatorname{ran} X_2$ (but not necessarily invariant under strictly increasing functions in general, see Example 7.22 below).
4) If (X_{i1}, X_{i2}) has copula C_i and $\mathbb{E}[X_{ij}^2] < \infty$, $i,j \in \{1,2\}$, then $C_1(u_1,u_2) \leq C_2(u_1,u_2)\ \forall\, u_1,u_2$ implies $\rho(X_{11}, X_{12}) \leq \rho(X_{21}, X_{22})$.

Proof.

1) Cauchy–Schwarz inequality $|\langle X, Y\rangle| \le \|X\|\|Y\|$ ($L^2 = \{\text{rv } X : \mathbb{E}[X^2] < \infty\}$ is a Hilbert space with inner product $\langle X, Y\rangle = \mathbb{E}[XY]$).

2) $\mathrm{Cov}[X_1, X_2] = \mathbb{E}[(X_1 - \mathbb{E}X_1)(X_2 - \mathbb{E}X_2)] \underset{\text{ind.}}{=} \mathbb{E}[X_1 - \mathbb{E}X_1]\mathbb{E}[X_2 - \mathbb{E}X_2] = 0 \Rightarrow \rho = 0$.

3) $\mathrm{Cov}[a_1X_1+b_1, a_2X_2+b_2] = a_1a_2\,\mathrm{Cov}[X_1, X_2]$ and $\sqrt{\mathrm{Var}[a_jX_j + b_j]} = \sqrt{a_j^2\,\mathrm{Var}\,X_j} = |a_j|\sqrt{\mathrm{Var}\,X_j}$, $j \in \{1, 2\}$, imply that for $a_1, a_2 > 0$,

$$\rho(a_1X_1 + b_1, a_2X_2 + b_2) = \frac{a_1a_2}{|a_1|\,|a_2|}\rho(X_1, X_2) = \rho(X_1, X_2).$$

4) $$\mathrm{Cov}[X_{11}, X_{12}] \underset{\text{Sklar}}{\overset{\text{Hoeff.}}{=}} \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\big(C_1(F_1(x_1), F_2(x_2)) - F_1(x_1)F_2(x_2)\big)\,dx_1dx_2$$
$$\le \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\big(C_2(F_1(x_1), F_2(x_2)) - F_1(x_1)F_2(x_2)\big)\,dx_1dx_2 = \mathrm{Cov}[X_{21}, X_{22}].$$

□

Drawbacks of linear correlation

1) ρ depends on the marginal distributions. In particular, for ρ to exist, $\operatorname{Var} X_1, \operatorname{Var} X_2$ must exist!

   Example 7.22 (ρ not invariant under general increasing functions)
   Consider $X_1, X_2 \overset{\text{ind.}}{\sim} \operatorname{Par}(3)$ (df $F(x) = 1 - x^{-3}$, $x \ge 1$; $\operatorname{Var} X_1, \operatorname{Var} X_2$ exist) $\Rightarrow$ $\rho(X_1, X_2) = 0$ but $\rho(X_1^2, X_2)$ does not even exist since $X_1^2 \sim \operatorname{Par}(3/2)$ and thus $\operatorname{Var}[X_1^2] = \infty$.

2) Linear correlation only gives a scalar summary of linear dependence. It is not invariant with respect to strictly increasing transformations in general, i.e.,

$$\rho(T_1(X_1), T_2(X_2)) \neq \rho(X_1, X_2).$$

   (assuming the left-hand side to exist; see Example 7.22).

3) X_1, X_2 independent $\Rightarrow \rho(X_1, X_2) = 0$. But $\rho(X_1, X_2) = 0 \not\Rightarrow X_1, X_2$ independent

Example 7.23 (Uncorrelated $\not\Rightarrow$ independent)

Consider the two risks

$$X_1 = Z \quad \text{(Profit \& Loss Country A)},$$
$$X_2 = VZ \quad \text{(Profit \& Loss Country B)},$$

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

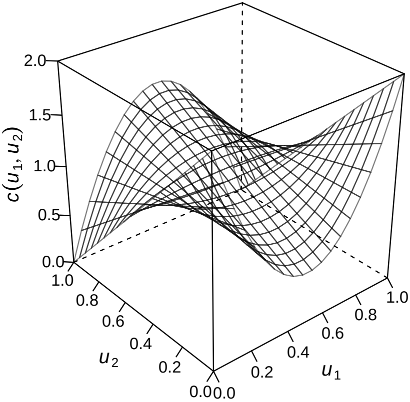
- Only known exception: $\mathrm{N}_2(\boldsymbol{\mu}, \Sigma)$ (uncorrelatedness $\underset{\text{density}}{\Rightarrow}$ independence)
- Compare Example 7.23 with $\boldsymbol{X}' = (X_1', X_2') \sim \mathrm{N}_2(\boldsymbol{0}, I_2)$. Both have $\mathrm{N}(0,1)$ margins and $\rho = 0$, but the copula of $\boldsymbol{X}'$ is the Π whereas the copula of $\boldsymbol{X} = (X_1, X_2)$ is $C(\boldsymbol{u}) = 0.5W(\boldsymbol{u}) + 0.5M(\boldsymbol{u}) \neq \Pi$. Thus F_1, F_2 and ρ do not uniquely determine H. Another example of this type is given by Fallacy 1 below.

Correlation fallacies

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

This is true for bivariate elliptical distributions, but wrong in general.

Counter-example: $C(u_1, u_2) = u_1 u_2 (1 - 2\theta(u_1 - \frac{1}{2})(u_1 - 1)(u_2 - 1))$



Properties

1) Take $F_1 = F_2 = \mathrm{U}[0,1]$ (extends to $\mathbb{E}[X_j^2] < \infty$ and F_1 symmetric about 0).
2) $\rho = 0$ for all $\theta \in [-1, 1]$ (see below)
3) There are ∞-many such models H (∞-many θ; plot: $\theta = 1$)

In particular, $\rho = 0 \not\Rightarrow C = \Pi$!

Reasoning for $\rho = 0$ for all $\theta \in [-1, 1]$:

By Hoeffding's identity and since $F_1 = F_2 = \mathrm{U}[0, 1]$ (for other margins, use Sklar's Theorem),

$$\rho = \frac{\mathrm{Cov}[X_1, X_2]}{\sqrt{\mathrm{Var}\, X_1 \,\mathrm{Var}\, X_2}} \underset{\text{Hoeff.}}{\overset{\mathrm{U}[0,1]}{=}} 12 \int_0^1 \int_0^1 (C(u_1, u_2) - u_1 u_2)\, du_1 du_2.$$

Now consider

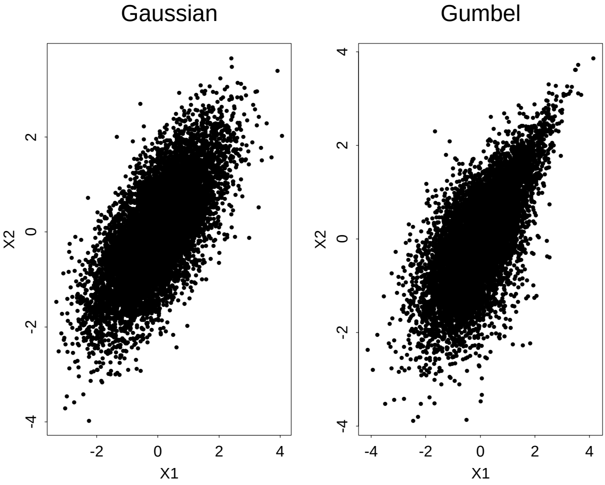
$$C(u_1, u_2) = u_1 u_2 + g_1(u_1) g_2(u_2)$$

with $g_j(0) = g_j(1) = 0$ and $g_1'(u_1) g_2'(u_2) \geq -1$ (check that C is a copula by computing its density). Then

$$\rho = 12 \int_0^1 g_1(u_1)\, du_1 \int_0^1 g_2(u_2)\, du_2$$

$\Rightarrow$ If g_1 is point symmetric about $1/2$, then $\rho = 0$.

Another counter-example: Gauss and Gumbel copulas compared



Margins are $\mathrm{N}(0, 1)$, $\rho = 0.7$; yet different multivariate models.

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

- This is true if (X_1, X_2) is elliptically distributed (we have seen that if $\mathbb{E}[R^2] < \infty$, $\mathrm{Cov}\,\boldsymbol{X} = \frac{\mathbb{E}[R^2]}{d}\Sigma$ and $\mathrm{Cor}\,\boldsymbol{X} = P$, the correlation matrix corresponding to the scale matrix Σ; see Example 6.21) but wrong in general; see below.
- **Reasoning 1:** If F_1 and F_2 are not of the same type, $\rho(X_1, X_2) < 1$; see Proposition 7.21 1).
- **Reasoning 2:** Hoeffding's identity

$$\mathrm{Cov}[X_1, X_2] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(C(F_1(x_1), F_2(x_2)) - F_1(x_1)F_2(x_2))\,dx_1 dx_2.$$

implies bounds on attainable ρ via

$$\rho_{\mathsf{min}} \le \rho \le \rho_{\mathsf{max}} \text{ (attained for } C = W \text{ and } C = M\text{, respectively).}$$

Example 7.24 (Bounds for a model with $\mathrm{LN}(0,\sigma_j^2)$ margins)

Let $X_j \sim \mathrm{LN}(0,\sigma_j^2)$, $j \in \{1,2\}$.

$C = W:$ $(X_1, X_2) = (\exp(\sigma_1 Z), \exp(-\sigma_2 Z))$ for $Z \sim \mathrm{N}(0,1)$. Then

$$\begin{aligned}\mathrm{Cov}[X_1, X_2] &= \mathbb{E}[X_1 X_2] - \mathbb{E}[X_1]\mathbb{E}[X_2] \\ &\underset{\mathbb{E}[\exp(tZ)]=e^{t^2/2}}{=} e^{(\sigma_1-\sigma_2)^2/2} - e^{(\sigma_1^2+\sigma_2^2)/2}.\end{aligned}$$

With $\mathrm{Var}[\mathrm{LN}(\mu,\sigma^2)] = (e^{\sigma^2}-1)e^{2\mu+\sigma^2}$, it follows that

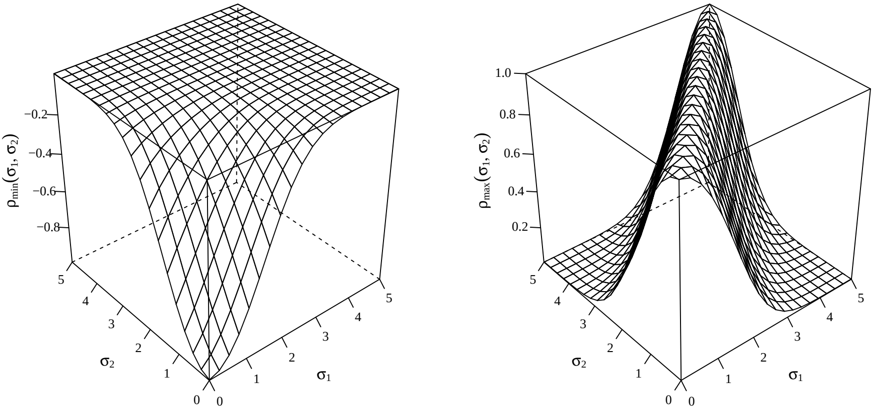
$$\rho_{\mathsf{min}} = \tilde{\rho}(\sigma_1, -\sigma_2),$$

where

$$\tilde{\rho}(\sigma_1,\sigma_2) = \frac{e^{(\sigma_1+\sigma_2)^2/2} - e^{(\sigma_1^2+\sigma_2^2)/2}}{\sqrt{(e^{\sigma_1^2}-1)e^{\sigma_1^2}}\sqrt{(e^{\sigma_2^2}-1)e^{\sigma_2^2}}}$$

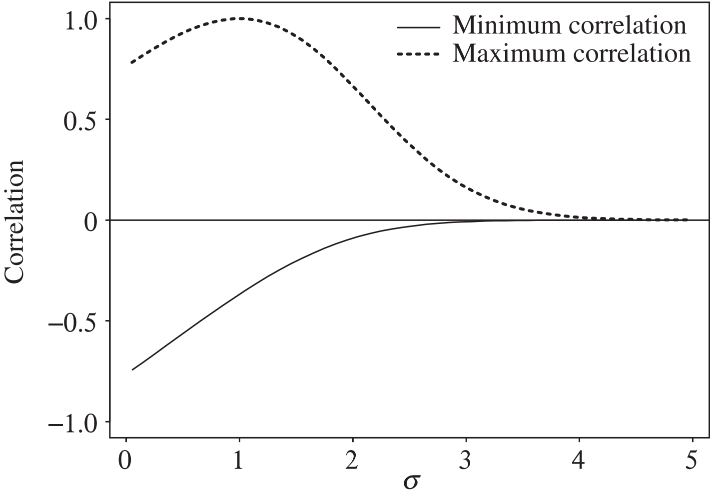
$C = M:$ Similarly, $\rho_{\mathsf{max}} = \tilde{\rho}(\sigma_1, \sigma_2)$.

A picture is worth a thousand words...

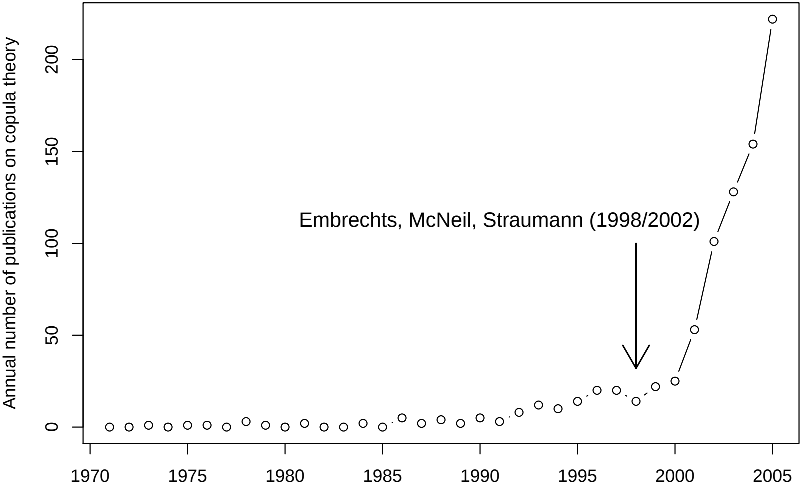


Example: For $\sigma_1^2 = 1$, $\sigma_2^2 = 16$ one has $\rho \in [-0.0003,\ 0.0137]$!

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



Note: This industry-inspired example has been influential...



The dataset stems from Genest et al. (2009).

Fallacy 3: ρ maximal (i.e., $C = M$) $\Rightarrow \mathrm{VaR}_\alpha(X_1 + X_2)$ 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, see Proposition 7.21 1), VaR_α is subadditive (so additivity provides the largest possible bound), and VaR_α is comonotone additive; see Propositions 6.24 and 7.19).
- Any superadditivity example $\mathrm{VaR}_\alpha(X_1+X_2) > \mathrm{VaR}_\alpha(X_1)+\mathrm{VaR}_\alpha(X_2)$ $(= \mathrm{VaR}_\alpha(X_1 + X_2)$ under maximal correlation, i.e., $C = M)$ serves as a counterexample; see Section 2.3.5.

7.2.3 Rank correlation

To overcome (some) of the deficiencies of ρ, Scarsini (1984) introduced:

Definition 7.25 (Rank correlation coefficient)

A measure of association $\kappa = \kappa(X_1, X_2) = \kappa(C)$ between two continuously distributed random variables X_1 and X_2 with copula C is a *rank correlation coefficient* if

1) κ exists for every pair (X_1, X_2) of cont. distributed random variables;
2) $-1 \leq \kappa \leq 1$, $\kappa(W) = -1$, and $\kappa(M) = 1$;
3) $\kappa(X_1, X_2) = \kappa(X_2, X_1)$;
4) X_1 and X_2 being independent implies $\kappa(X_1, X_2) = \kappa(\Pi) = 0$;
5) $\kappa(-X_1, X_2) = -\kappa(X_1, X_2)$;
6) $C_1(\boldsymbol{u}) \leq C_2(\boldsymbol{u})$ for all $\boldsymbol{u} \in [0,1]^2$ implies $\kappa(C_1) \leq \kappa(C_2)$;
7) $C_n \to C$ $(n \to \infty)$ pointwise implies $\lim_{n\to\infty} \kappa(C_n) = \kappa(C)$.

Proposition 7.26 (Basic properties of κ)

Let κ be a rank correlation coefficient for two continuously distributed random variables $X_1 \sim F_1$ and $X_2 \sim F_2$. Then

1) $\kappa(X_1, X_2) = \kappa(C)$ (κ only depends on C).
2) if T_j is a strictly increasing function on $\operatorname{ran} X_j$, $j \in \{1, 2\}$, then $\kappa(T_1(X_1), T_2(X_2)) = \kappa(X_1, X_2)$.

Proof.

1) Set $(U_1, U_2) = (F_1(X_1), F_2(X_2))$. By the invariance principle, (X_1, X_2) and (U_1, U_2) have the same copula C. Thus, by 6), $\kappa(U_1, U_2) \leq \kappa(X_1, X_2)$, but also $\kappa(X_1, X_2) \leq \kappa(U_1, U_2)$, so $\kappa(X_1, X_2) = \kappa(U_1, U_2)$ ($\Rightarrow$ only depends on C).
2) Invariance principle $\Rightarrow$ The copula C of (X_1, X_2) equals the copula of $(T_1(X_1), T_2(X_2))$. Hence $\kappa(T_1(X_1), T_2(X_2)) = \kappa(C) = \kappa(X_1, X_2)$.

□

Rank correlation coefficients are…

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

Examples: Kendall's tau and Spearman's rho

Definition 7.27 (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

$$\begin{aligned}\tau &= \mathbb{E}[\operatorname{sign}((X_1 - X_1')(X_2 - X_2'))] \\ &= \mathbb{P}((X_1 - X_1')(X_2 - X_2') > 0) - \mathbb{P}((X_1 - X_1')(X_2 - X_2') < 0),\end{aligned}$$

where $\operatorname{sign}(x) = \mathbb{1}_{(0,\infty)}(x) - \mathbb{1}_{(-\infty,0)}(x)$.

Proposition 7.28 (Formula for Kendall's tau)

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

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

Proof. Let (X_1', X_2') be an independent copy of (X_1, X_2). Then

$$\begin{aligned}
\tau &= \mathbb{P}((X_1 - X_1')(X_2 - X_2') > 0) - \mathbb{P}((X_1 - X_1')(X_2 - X_2') < 0) \\
&= 2\,\underbrace{\mathbb{P}((X_1 - X_1')(X_2 - X_2') > 0)}_{=2\mathbb{P}(X_1 < X_1', X_2 < X_2')} - 1 = 4\mathbb{P}(U_1 \le U_1', U_2 \le U_2') - 1 \\
&= 4 \int_0^1 \int_0^1 \mathbb{P}(U_1 \le u_1, U_2 \le u_2)\, dC(u_1, u_2) - 1
\end{aligned}$$

□

- For computing τ, $\int_{[0,1]^2} C(\boldsymbol{u})\, d\tilde{C}(\boldsymbol{u}) = \frac{1}{2} - \int_{[0,1]^2} D_1 C(\boldsymbol{u}) D_2 \tilde{C}(\boldsymbol{u})\, d\boldsymbol{u}$ is often helpful; see Li et al. (2002).

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

$$\hat{\tau}_n = \frac{1}{\binom{n}{2}} \sum_{1\le i_1<i_2\le n} \operatorname{sign}((X_{i_1 1} - X_{i_2 1})(X_{i_1 2} - X_{i_2 2})). \qquad (38)$$

- This is often used to compute estimators for P (the correlation matrix corresponding to the scale matrix Σ) for elliptical copulas:
  1) Compute $\hat{\tau}_{n,j_1 j_2}$ for all $1 \le j_1 < j_2 \le d$ (column indices of all pairs).
  2) Invert (32) to get $\hat{P}'_{j_1 j_2} = \sin(\frac{\pi}{2}\tau)$, $j_1 \neq j_2$ (see later why).
  3) Use, e.g., `nearPD(, corr=TRUE)` in the R package `Matrix` to get a positive-definite matrix $\hat{P}$ close to $\hat{P}'$.

Definition 7.29 (Spearman's rho)
Let $X_j \sim F_j$ with F_j continuous, $j \in \{1,2\}$. *Spearman's rho* is defined by $\rho_{\mathsf{S}} = \rho(F_1(X_1), F_2(X_2))$.

Proposition 7.30 (Formula for Spearman's rho)

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

$$\rho_{\mathsf{S}} = 12 \int_0^1 \int_0^1 C(u_1, u_2)\, du_1 du_2 - 3.$$

Proof. By Hoeffding's identity, we have $\rho_{\mathsf{S}}(X_1, X_2) = \rho(F_1(X_1), F_2(X_2))$ $= 12 \int_0^1 \int_0^1 (C(u_1, u_2) - u_1 u_2)\, du_1 du_2 = 12 \int_0^1 \int_0^1 C(u_1, u_2)\, du_1 du_2 - 3.$ □

- An estimator $\hat{\rho}_{\mathsf{S},n}$ is given by the sample correlation computed from pseudo-observations (marginal empirical dfs applied componentwise).
- For κ being either Spearman's rho or Kendall's tau, Embrechts et al. (2002) show that $\kappa = \pm 1$ if and only if X_1, X_2 are co-/countermonotonic.
- Fallacy 2 (For F_1, F_2, any $\rho \in [-1, 1]$ is attainable) is solved. Take

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

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

- Fallacy 1 (F_1, F_2, ρ uniquely determine H) 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 3 ($C = M$ implies $\mathrm{VaR}_\alpha(X_1+X_2)$ 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 $\mathrm{VaR}_\alpha(X_1 + X_2)$; see our superadditivity examples.
- Note that for $\boldsymbol{S} \sim \mathrm{U}(\{\boldsymbol{x} \in \mathbb{R}^2 : \|\boldsymbol{x}\| = 1\})$, $S_1 \stackrel{\text{d}}{=} -S_1$, so $\kappa(S_1, S_2) = \kappa(-S_1, S_2) = -\kappa(S_1, S_2)$ and hence $\kappa(S_1, S_2) = 0$ although (S_1, S_2) are clearly not independent.
- Nevertheless, rank correlations are useful to overall summarize dependence or parameterize copula families to make dependence comparable (they can obviously not replace the underlying copula, though).

7.2.4 Coefficients of tail dependence

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

Definition 7.31 (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

$$\lambda_L = \lim_{u \downarrow 0} \mathbb{P}(X_2 \le F_2^-(u) \,|\, X_1 \le F_1^-(u)),$$

$$\lambda_U = \lim_{u \uparrow 1} \mathbb{P}(X_2 > F_2^-(u) \,|\, X_1 > F_1^-(u)).$$

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 (note that $F_j \uparrow$ on $\operatorname{ran} X_j$)

$$\mathbb{P}(X_2 \le F_2^-(u) \,|\, X_1 \le F_1^-(u)) = \frac{\mathbb{P}(X_2 \le F_2^-(u), X_1 \le F_1^-(u))}{\mathbb{P}(X_1 \le F_1^-(u))}$$

$$= \frac{\mathbb{P}(F_2(X_2) \le F_2(F_2^-(u)), F_1(X_1) \le F_1(F_1^-(u)))}{\mathbb{P}(F_1(X_1) \le F_1(F_1^-(u)))}$$

$$\underset{\text{(GI4)}}{=} \frac{\mathbb{P}(F_1(X_1) \le u, F_2(X_2) \le u)}{\mathbb{P}(F_1(X_1) \le u)} = \frac{C(u,u)}{u}, \quad u \in (0,1),$$

  so $\lambda_L = \lim_{u \downarrow 0} \frac{C(u,u)}{u}$.
- If $u \mapsto C(u,u)$ is differentiable in a neighborhood of 0 and the limit exists, then $\lambda_L = \lim_{u\downarrow 0} \frac{d}{du} C(u,u)$ (l'Hôpital's Rule).
- If C is totally differentiable in a neighborhood of 0 and the limit exists, then $\lambda_L = \lim_{u\downarrow 0}(\mathrm{D}_1\, C(u,u) + \mathrm{D}_2\, C(u,u))$ (Chain Rule). If C is symmetric, then $\lambda_L = 2 \lim_{u\downarrow 0} \mathrm{D}_1\, C(u,u)$.

- Similarly as above, for the upper tail-dependence coefficient,
$$\lambda_U = \lim_{u\uparrow 1} \frac{1-2u+C(u,u)}{1-u} = \lim_{u\downarrow 0} \frac{\hat{C}(u,u)}{u}$$
Also, $\lambda_U = 2 - \lim_{u\uparrow 1} \frac{d}{du} C(u,u)$ and $\lambda_U = \lim_{u\uparrow 1}(\mathrm{D}_1\, C(u,u) + \mathrm{D}_2\, C(u,u))$ if the corresponding limits exist.
- λ_L, λ_U not necessarily exist, but they do in all practical cases.
- For elliptical copulas, $\lambda_L = \lambda_U =: \lambda$ (true for all radially symmetric copulas); for C_P^{Ga}, $\lambda = 0$, and for $C_{\nu,P}^t$, $\lambda > 0$ (see later).
- For Archimedean copulas with strict ψ, a substitution and l'Hôpital's Rule show:
$$\lambda_L = \lim_{u\downarrow 0} \frac{\psi(2\psi^{-1}(u))}{u} = \lim_{t\to\infty} \frac{\psi(2t)}{\psi(t)} = 2 \lim_{t\to\infty} \frac{\psi'(2t)}{\psi'(t)},$$
$$\lambda_U = 2 - \lim_{u\uparrow 1} \frac{1-\psi(2\psi^{-1}(u))}{1-u} = 2 - \lim_{t\downarrow 0} \frac{1-\psi(2t)}{1-\psi(t)} = 2 - 2 \lim_{t\downarrow 0} \frac{\psi'(2t)}{\psi'(t)}$$
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 $\boldsymbol{X} \overset{\text{d}}{=} \boldsymbol{\mu} + \sqrt{W} A \boldsymbol{Z}$ ($\boldsymbol{X} \overset{\text{d}}{=} \boldsymbol{m}(W) + \sqrt{W} A \boldsymbol{Z}$); e.g., Gauss, t copulas.

7.3.1 Tail dependence

Coefficients of tail dependence

Let $(U_1, U_2) \sim C$ for a normal variance mixture copula C. Then

$$\lambda \underset{\text{symm.}}{\overset{\text{radial}}{=}} \lambda_L \underset{\text{symm.}}{=} 2 \lim_{q \downarrow 0} \mathrm{D}_1 C(q, q) \underset{\text{Th. 7.16}}{=} 2 \lim_{q \downarrow 0} \mathbb{P}(U_2 \le q \,|\, U_1 = q).$$

Example 7.32 (λ for the Gauss and t copula)

- This result goes back to Sibuya (1959). Let $\boldsymbol{U} = (\Phi(X_1), \Phi(X_2))$ with $\boldsymbol{X} \sim \mathrm{N}_2(\boldsymbol{0}, P)$, where $P = \left(\begin{smallmatrix} 1 & \rho \\ \rho & 1 \end{smallmatrix}\right)$, i.e., $\boldsymbol{U} \sim C_P^{\text{Ga}}$. Recall that $X_2 \,|\, X_1 = x \sim \mathrm{N}(\rho x, 1 - \rho^2)$. This implies that $\lambda = 2 \lim_{x \downarrow -\infty} \mathbb{P}(X_2 \le x \,|\, X_1 = x) = 2 \lim_{x \downarrow -\infty} \Phi\left(\frac{x(1-\rho)}{\sqrt{1-\rho^2}}\right) = \mathbb{1}_{\{\rho = 1\}}$ (essentiallly no tail dependence).

- For $C^t_{\nu,P}$, one can show that (via $f_{X_2|X_1}(x_2\,|\,x_1)=\frac{f_{X_1,X_2}(x_1,x_2)}{f_{X_1}(x_1)}$)

$$X_2\,|\,X_1=x\sim t_{\nu+1}\Bigl(\rho x,\frac{(1-\rho^2)(\nu+x^2)}{\nu+1}\Bigr)$$

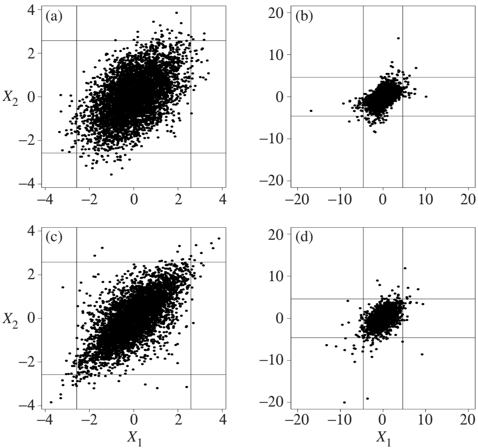
and thus $\mathbb{P}(X_2\le x\,|\,X_1=x)=t_{\nu+1}\Bigl(\frac{x-\rho x}{\sqrt{\frac{(1-\rho^2)(\nu+x^2)}{\nu+1}}}\Bigr)$. Hence

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

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) $\mathrm{N}_2(\mathbf{0}, P = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix})$, $\rho = 0.5$;

(b) C^{Ga}_{ρ} with t_4 margins (same dependence as in (a));

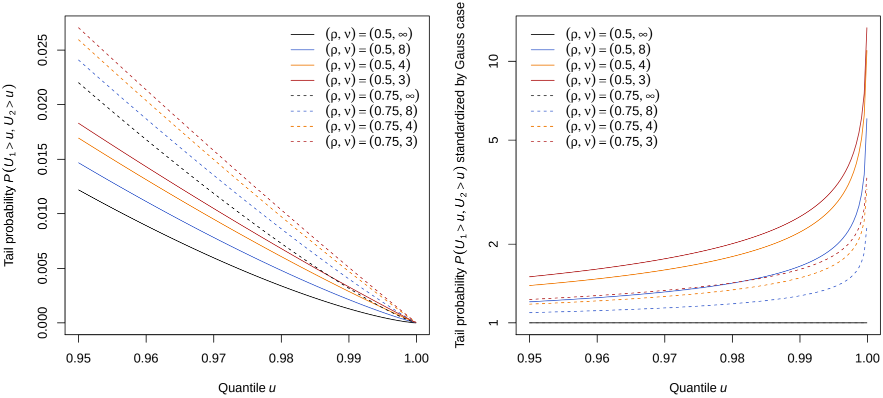
(c) $C^{t}_{4,\rho}$ with $\mathrm{N}(0,1)$ margins;

(d) $t_2(4, \mathbf{0}, P)$ (same dependence as in (c)).

Lines denote 0.005- and 0.995-quantiles.

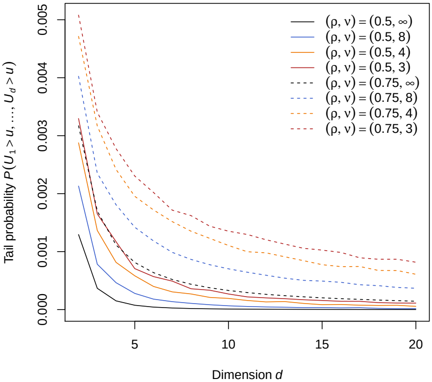
Note the different number of points in the bivariate tails (all models have the same Kendall's tau!)

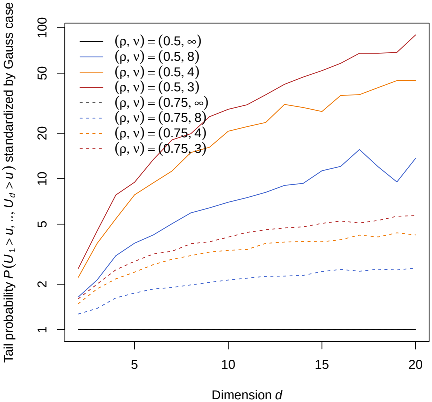
Joint tail probabilities $\mathbb{P}(U_1 > u, U_2 > u)$ for $d = 2$



- **Left:** The higher ρ or the smaller ν, the larger $\mathbb{P}(U_1 > u, U_2 > u)$.
- **Right:** $u \mapsto \dfrac{\mathbb{P}(U_1 > u, U_2 > u)}{\mathbb{P}(V_1 > u, V_2 > u)} \overset{\text{radial}}{\underset{\text{symm.}}{=}} \dfrac{C^t_{\nu,\rho}(u,u)}{C^{\mathrm{Ga}}_{\rho}(u,u)}$

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:** $d \mapsto \dfrac{\mathbb{P}(U_1 > u, \dots, U_d > u)}{\mathbb{P}(V_1 > u, \dots, V_d > u)} \underset{\text{symm.}}{\overset{\text{radial}}{=}} \dfrac{C^t_{\nu,\rho}(u, \dots, u)}{C^{\text{Ga}}_{\rho}(u, \dots, u)}$ for $u = 0.99$.

Example 7.33 (Joint tail probabilities: an interpretation)

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

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

  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 $\boldsymbol{X}$ is $C^{t}_{\nu=4,\rho=0.5}$, 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!

Felix Salmon: "Recipe for Disaster: The Formula That Killed Wall Street"

$$\Pr[T_A<1, T_B<1] = \phi_2(\phi^{-1}(F_A(1)), \phi^{-1}(F_B(1)), \gamma)$$

Here's what killed your 401(k) *David X. Li's Gaussian copula function as first published in 2000. Investors exploited it as a quick—and fatally flawed—way to assess risk. A shorter version appears on this month's cover of* Wired.

Probability
Specifically, this is a joint default probability—the likelihood that any two members of the pool (A and B) will both default. It's what investors are looking for, and the rest of the formula provides the answer.

Survival times
The amount of time between now and when A and B can be expected to default. Li took the idea from a concept in actuarial science that charts what happens to someone's life expectancy when their spouse dies.

Equality
A dangerously precise concept, since it leaves no room for error. Clean equations help both quants and their managers forget that the real world contains a surprising amount of uncertainty, fuzziness, and precariousness.

Copula
This couples (hence the Latinate term copula) the individual probabilities associated with A and B to come up with a single number. Errors here massively increase the risk of the whole equation blowing up.

Distribution functions
The probabilities of how long A and B are likely to survive. Since these are not certainties, they can be dangerous: Small miscalculations may leave you facing much more risk than the formula indicates.

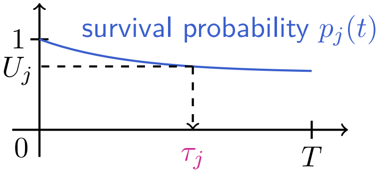
Gamma
The all-powerful correlation parameter, which reduces correlation to a single constant—something that should be highly improbable, if not impossible. This is the magic number that made Li's copula function irresistible.

Application to credit risk

Intensity-based default model:

$$p_j(t) = \exp\Big(-\int_0^t \lambda_j(s)\,ds\Big)$$

$$\tau_j = \inf\{t \ge 0 : p_j(t) \le U_j\}$$



Note: $\lambda_U = 0$ (as for the Gauss copula!)

$\Rightarrow$ (Almost) no joint defaults! (p_j typically very flat)

Copulas for the triggers $\boldsymbol{U}$:

1) Li (2000): Gauss (Sibuya (1959): $\lambda_U = 0$)

2) Schönbucher and Schubert (2001): Archimedean ($\lambda_U > 0$)

3) Hofert and Scherer (2011): nested Archimedean ($\lambda_U > 0$, hierarchies)

Typical application: CDO pricing models based on iTraxx data.

7.3.2 Rank correlations

Lemma 7.34

Let $\boldsymbol{X} \sim \mathrm{E}_2(\mathbf{0}, \Sigma, \psi)$ with $\mathbb{P}(\boldsymbol{X} = \mathbf{0}) = 0$ and $\rho = P_{12} = \mathrm{Cor}[\Sigma]_{12}$. Then

$$\mathbb{P}(X_1 > 0, X_2 > 0) = \frac{1}{4} + \frac{\arcsin\rho}{2\pi}.$$

Proof.

- Note that $\boldsymbol{Y} = \left(\begin{smallmatrix} 1/\sqrt{\sigma_{11}} & 0 \\ 0 & 1/\sqrt{\sigma_{22}} \end{smallmatrix}\right)\boldsymbol{X} \sim \mathrm{E}_2(\mathbf{0}, P, \psi)$ with $P = \left(\begin{smallmatrix} 1 & \rho \\ \rho & 1 \end{smallmatrix}\right)$.
- Let $A = \left(\begin{smallmatrix} 1 & 0 \\ \rho & \sqrt{1-\rho^2} \end{smallmatrix}\right)$ so that $AA^\top = P$. Then $\boldsymbol{Y} \overset{\mathrm{d}}{=} RA\boldsymbol{U} \overset{\mathrm{d}}{=} RA\left(\begin{smallmatrix} \cos\Theta \\ \sin\Theta \end{smallmatrix}\right)$, $\Theta \sim \mathrm{U}[-\pi, \pi)$ independent of R.
- With $\varphi = \arcsin\rho$, we have $\boldsymbol{Y} \overset{\mathrm{d}}{=} R\left(\begin{smallmatrix} \cos\Theta \\ \sin\varphi\cos\Theta + \cos\varphi\sin\Theta \end{smallmatrix}\right) = \left(\begin{smallmatrix} \cos\Theta \\ \sin(\varphi+\Theta) \end{smallmatrix}\right)$.
- Thus $\mathbb{P}(X_1 > 0, X_2 > 0) = \mathbb{P}(Y_1 > 0, Y_2 > 0) = \mathbb{P}(\cos\Theta > 0, \sin(\varphi + \Theta) > 0) = \mathbb{P}(\Theta \in (-\frac{\pi}{2}, \frac{\pi}{2}), \varphi + \Theta \in (0, \pi) = \mathbb{P}(\Theta \in (-\frac{\pi}{2}, \frac{\pi}{2}), \Theta \in (-\varphi, \pi - \varphi) = \mathbb{P}(\Theta \in (-\varphi, \frac{\pi}{2}]) = (\frac{\pi}{2} - (-\varphi))/(2\pi)$. $\square$

Lemma 7.35 (Representation of Spearman's rho)
Let $(U_1, U_2) \sim C$ and $\tilde{U}_1, \bar{U}_2 \overset{\text{ind.}}{\sim} \mathrm{U}[0,1]$ be independent. Then $\rho_{\mathsf{S}} = \rho_{\mathsf{S}}(U_1, U_2) = 12\mathbb{P}(U_1 \le \tilde{U}_1, U_2 \le \bar{U}_2) - 3$.

Proof. $12\mathbb{P}(U_1 \le \tilde{U}_1, U_2 \le \bar{U}_2) - 3 = 12\mathbb{E}[\mathbb{P}(\tilde{U}_1 > U_1, \bar{U}_2 > U_2 \,|\, U_1, U_2)] - 3 \underset{\text{ind.}}{=} 12\mathbb{E}[(1-U_1)(1-U_2)] - 3 = 12\mathbb{E}[U_1 U_2] - 3 = \rho_{\mathsf{S}}(U_1, U_2)$. □

Theorem 7.36 (Spearman's rho for Gauss copulas)
Let $(U_1, U_2) \sim C^{\mathrm{Ga}}_{\rho}$. Then $\rho_{\mathsf{S}} = \rho_{\mathsf{S}}(U_1, U_2) = \frac{6}{\pi} \arcsin \frac{\rho}{2}$.

Proof. Let $(X_1, X_2) \sim \mathrm{N}_2(\mathbf{0}, P)$ with $P_{12} = \rho$, indep. of $\tilde{X}_1, \bar{X}_2 \overset{\text{ind.}}{\sim} \mathrm{N}(0,1)$. With $\boldsymbol{Y} = (\tilde{X}_1 - X_1, \bar{X}_2 - X_2) \sim \mathrm{N}_2(\mathbf{0}, I_2 + P)$ $(\mathrm{Cor}[Y_1, Y_2] = \rho/2)$

$$\rho_{\mathsf{S}} \underset{\Phi^{-1}}{\overset{\text{L. 7.35}}{=}} 12\mathbb{P}(X_1 \le \tilde{X}_1, X_2 \le \bar{X}_2) - 3 = 12\mathbb{P}(Y_1 \ge 0, Y_2 \ge 0) - 3$$

$$\underset{\text{L.7.34}}{\overset{\text{cont.}}{=}} 12\Big(\frac{1}{4} + \frac{\arcsin(\rho/2)}{2\pi}\Big) - 3 = \frac{6}{\pi} \arcsin \frac{\rho}{2}.$$ □

Proposition 7.37 (Spearman's rho for normal variance mixtures)
Let $\boldsymbol{X} \sim M_2(\mathbf{0}, P, \hat{H})$ with $\mathbb{P}(\boldsymbol{X} = \mathbf{0}) = 0$, $\rho = P_{12}$. Then

$$\rho_{\mathsf{S}} = \frac{6}{\pi}\mathbb{E}\Big[\arcsin \frac{W\rho}{\sqrt{(W+\tilde{W})(W+\bar{W})}}\Big],$$

for $W, \tilde{W}, \bar{W} \overset{\text{ind.}}{\sim} H$ with Laplace–Stieltjes transform $\hat{H}$.

Proof. $\boldsymbol{X} \overset{\mathrm{d}}{=} \sqrt{W}\boldsymbol{Z}$ for $\boldsymbol{Z} \sim \mathrm{N}_2(\mathbf{0}, P)$. Let $\tilde{Z}, \bar{Z} \sim \mathrm{N}(0,1)$ and assume $\boldsymbol{Z}$, $\tilde{Z}$, $\bar{Z}$, W, $\tilde{W}$ and $\bar{W}$ are all independent. Let

$$\tilde{X} = \sqrt{\tilde{W}}\tilde{Z}, \quad \bar{X} = \sqrt{\bar{W}}\bar{Z},$$

$$Y_1 = X_1 - \tilde{X} = \sqrt{W}Z_1 - \sqrt{\tilde{W}}\tilde{Z},$$

$$Y_2 = X_2 - \bar{X} = \sqrt{W}Z_2 - \sqrt{\bar{W}}\bar{Z}.$$

$$
\begin{aligned}
\rho_{\mathsf{S}}(X_1, X_2) &\underset{\Phi^{-1}}{\overset{\text{L. 7.35}}{=}} 12\mathbb{P}(X_1 \le \tilde{X}_1, X_2 \le \bar{X}_2) - 3 \\
&= 6\mathbb{P}((X_1 - \tilde{X}_1)(X_2 - \bar{X}_2) > 0) - 3 \\
&= 3(2\mathbb{E}[\mathbb{P}(Y_1 Y_2 > 0 \,|\, W, \tilde{W}, \bar{W})] - 1) \\
&= 3(4\mathbb{E}[\mathbb{P}(Y_1 > 0, Y_2 > 0 \,|\, W, \tilde{W}, \bar{W})] - 1).
\end{aligned}
$$

Now note that $\boldsymbol{Y} \,|\, W, \tilde{W}, \bar{W} \sim \mathrm{N}_2(\mathbf{0}, \left(\begin{smallmatrix} W+\tilde{W} & W\rho \\ W\rho & W+\bar{W} \end{smallmatrix}\right))$ with $\rho(Y_1, Y_2)$ $= \frac{W\rho}{\sqrt{(W+\tilde{W})(W+\bar{W})}}$. Apply Lemma 7.34 to see that this equals

$$
\rho_{\mathsf{S}}(X_1, X_2) = 3\Big(4\mathbb{E}\Big[\frac{1}{4} + \frac{\arcsin \rho}{2\pi}\Big] - 1\Big) = \frac{12}{2\pi}\mathbb{E}[\arcsin \rho(Y_1, Y_2)]. \quad \square
$$

Proposition 7.38 (Kendall's tau for elliptical distributions)
Let $\boldsymbol{X} \sim \mathrm{E}_2(\mathbf{0}, P, \psi)$ with $\mathbb{P}(\boldsymbol{X}=\mathbf{0})=0$, $\rho=P_{12}$. Then $\tau = \frac{2}{\pi}\arcsin\rho$.

Proof.

- Let (X_1', X_2') be an independent copy of (X_1, X_2). We have already seen that $\tau = 2\mathbb{P}((X_1 - X_1')(X_2 - X_2') > 0) - 1$.
- With $\boldsymbol{X} \overset{\mathrm{d}}{=} RA\boldsymbol{U}$ and $\boldsymbol{X}' \overset{\mathrm{d}}{=} R'A\boldsymbol{U}' (\overset{\mathrm{d}}{=} -\boldsymbol{X}')$ we have $\boldsymbol{Y} = \boldsymbol{X} - \boldsymbol{X}' \overset{\mathrm{d}}{=} \mathbf{0} + A(R\boldsymbol{U} - R'\boldsymbol{U}')$. Note that the characteristic function of $-\boldsymbol{X}'$ is $\phi_{-\boldsymbol{X}'}(\boldsymbol{t}) = \phi_{\boldsymbol{X}'}(\boldsymbol{t}) = \phi_{\boldsymbol{X}}(\boldsymbol{t})$ so that $\phi_{\boldsymbol{Y}}(\boldsymbol{t}) \underset{\text{ind.}}{=} \phi_{\boldsymbol{X}}(\boldsymbol{t})\phi_{-\boldsymbol{X}'}(\boldsymbol{t}) = \phi_{\boldsymbol{X}}(\boldsymbol{t})^2$, hence $\boldsymbol{Y} \sim \mathrm{E}_2(\mathbf{0}, P, \psi^2)$.
- We thus obtain that

$$\begin{aligned}\tau &= 2\mathbb{P}(Y_1Y_2 > 0) - 1 = 2(\mathbb{P}(Y_1 > 0, Y_2 > 0) + \mathbb{P}(Y_1 < 0, Y_2 < 0)) - 1\\ &= 4\mathbb{P}(\boldsymbol{Y} > \mathbf{0}) - 1 \underset{\text{L.7.34}}{\overset{\text{cont.}}{=}} \frac{2}{\pi}\arcsin\rho. \end{aligned}$$

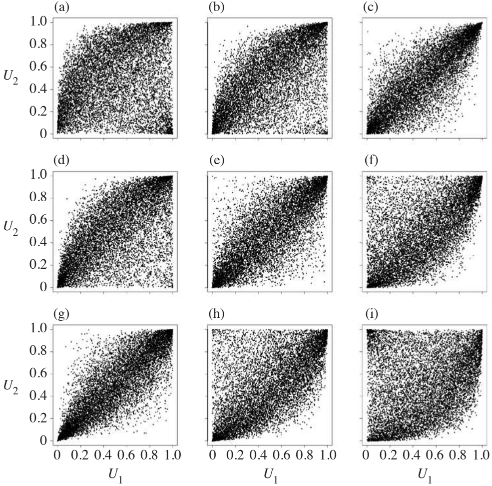
□

For a generalization to componentwise n.d. $\boldsymbol{X}$, see Lindskog et al. (2002).

7.3.3 Skewed normal mixture copulas

- ... are the copulas of normal mixture distributions which are not elliptical, e.g., the *skewed t copula* $C^t_{\nu,P,\boldsymbol{\gamma}}$ is the copula of $\boldsymbol{X} \sim \mathrm{GH}_d(-\frac{\nu}{2}, \nu, 0, \boldsymbol{\mu}, \Sigma, \boldsymbol{\gamma})$.
- The main advantage of $C^t_{\nu,P,\boldsymbol{\gamma}}$ over $C^t_{\nu,P}$ is its radial asymmetry (e.g., for modeling $\lambda_L \neq \lambda_U$)
- $C^t_{\nu,P,\boldsymbol{\gamma}}$ can be sampled as other implicit copulas; see Algorithm 7.12 (the evaluation of the margins requires numerical integration of a skewed t density; note that $X_j \sim \mathrm{GH}_1(-\frac{\nu}{2}, \nu, 0, \mu_j, \Sigma_{jj}, \gamma_j)$, $j \in \{1, \dots, d\}$).

10 000 samples from $C^t_{\nu=5,\ \rho=0.8,\ \gamma=0.8(\mathbb{1}_{\{i<2\}}-\mathbb{1}_{\{i>2\}},\mathbb{1}_{\{j>2\}}-\mathbb{1}_{\{j<2\}})}$



(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

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

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

  for $(W_1, \dots, W_S) \sim M(\mathrm{IG}(\frac{\nu_1}{2}, \frac{\nu_1}{2}), \dots, \mathrm{IG}(\frac{\nu_S}{2}, \frac{\nu_S}{2}))$ and $\boldsymbol{Y} \sim \mathrm{N}_d(\boldsymbol{0}, P)$ (so $\boldsymbol{Y} \overset{\mathrm{d}}{=} A\boldsymbol{Z}$ as before); see Demarta and McNeil (2005) for more details.

- Clearly, the marginals are t distributed, hence

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

  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

- *Archimedean copulas* are explicit copulas, arising from a construction principle of copulas directly (conditional independence approach).
- **Recall:** An (Archimedean) generator ψ is a function $\psi : [0,\infty) \to [0,1]$ which is $\downarrow$ on $[0, \inf\{t : \psi(t) = 0\}]$ and satisfies $\psi(0) = 1$, $\psi(\infty) = \lim_{t\to\infty} \psi(t) = 0$; the set of all generators is denoted by Ψ.

7.4.1 Bivariate Archimedean copulas

Theorem 7.39 (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.

Proof. See Nelsen (1999, pp. 91). □

- **Recall:** If ψ is strict, $\lambda_L = 2\lim_{t\to\infty} \frac{\psi'(2t)}{\psi'(t)}$ and $\lambda_U = 2 - 2\lim_{t\downarrow 0} \frac{\psi'(2t)}{\psi'(t)}$.

Proposition 7.40 (Kendall's tau for Archimedean copulas)
For a strict and twice-continuously differentiable ψ,

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

Proof. See Nelsen (1999, pp. 130). □

Example 7.41 (Outer power Clayton copula)
An *outer power Clayton copula* is generated by $\tilde{\psi}(t) = \psi(t^{1/\beta})$, $\beta \in [1,\infty)$, where $\psi(t) = (1+t)^{-1/\theta}$, $\theta \in (0,\infty)$, denotes the Clayton generator (with $\tau = \theta/(\theta+2)$). The quantities corresponding to $\tilde{\psi}$ are

$$\tilde{\tau} = 1 - \frac{1-\tau}{\beta} = 1 - \frac{2}{\beta(\theta+2)}, \quad \tilde{\lambda}_L = 2^{-1/(\theta\beta)}, \quad \tilde{\lambda}_U = 2 - 2^{1/\beta}.$$

The most widely used one-parameter Archimedean copulas are (see the R package `copula`):

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

Family	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 7.42 (Kimberling (1974))

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

Proof. See Kimberling (1974) or Hofert (2010, p. 54). □

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

Theorem 7.43 (Bernstein (1928))

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

Proof. See Feller (1971, pp. 439). □

We thus use the notation $\psi(t) = \mathcal{LS}[F](t)$ or $F(x) = \mathcal{LS}^{-1}[\psi](x)$.

Proposition 7.44 (Stochastic representation, related properties)
Let $\psi \in \Psi_\infty$ with $V \sim F = \mathcal{LS}^{-1}[\psi]$ and $E_1, \dots, E_d \overset{\text{ind.}}{\sim} \operatorname{Exp}(1)$ independent of V. Then

1) The survival copula of $\boldsymbol{X} = (E_1/V, \dots, E_d/V)$ is Archimedean with generator ψ.
2) $\boldsymbol{U} = (\psi(X_1), \dots, \psi(X_d)) \sim C$
3) The components of $\boldsymbol{U}$ are conditionally independent given V with conditional df $\mathbb{P}(U_j \le u \,|\, V = v) = \exp(-v\psi^{-1}(u))$.

Proof.

1) The joint survival function of $\boldsymbol{X}$ is given by

$$\bar{H}(\boldsymbol{x}) = \mathbb{P}(X_j > x_j \;\forall\, j) = \int_0^\infty \mathbb{P}(E_j/V > x_j \;\forall\, j \,|\, V = v)\, dF(v)$$
$$= \int_0^\infty \mathbb{P}(E_j > v x_j \;\forall\, j)\, dF(v) = \int_0^\infty \prod_{j=1}^d \exp(-v x_j)\, dF(v)$$

$$= \int_0^\infty \exp\Big(-v\sum_{j=1}^d x_j\Big)\,dF(v) = \psi\Big(\sum_{j=1}^d x_j\Big).$$

The jth 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 $\hat{C}(\boldsymbol{u}) = \bar{H}(\bar{F}_1^-(u_1), \dots, \bar{F}_d^-(u_d)) = \psi(\sum_{j=1}^d \psi^{-1}(u_j))$.

2) $\mathbb{P}(\boldsymbol{U} \le \boldsymbol{u}) = \mathbb{P}(X_j > \psi^{-1}(u_j)\ \forall\, j) \underset{1)}{=} \psi(\sum_{j=1}^d \psi^{-1}(u_j))$.

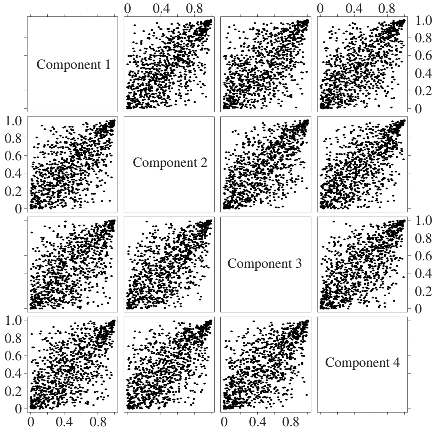
3) Cond. indep. is clear. Furthermore, $\mathbb{P}(U_j \le u \,|\, V = v) = \mathbb{P}(X_j > \psi^{-1}(u) \,|\, V = v) = \mathbb{P}(E_j > v\psi^{-1}(u)) = \exp(-v\psi^{-1}(u))$. □

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

Algorithm 7.45 (Marshall and Olkin (1988))

1) Sample $V \sim F = \mathcal{LS}^{-1}[\psi]$ (df corresponding to ψ; see tables above).
2) Sample $E_1, \dots, E_d \overset{\text{ind.}}{\sim} \mathrm{Exp}(1)$ independently of V.
3) Return $\boldsymbol{U} = (\psi(E_1/V), \dots, \psi(E_d/V))$ (conditional independence).

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



- 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 $\boldsymbol{X}_1, \dots, \boldsymbol{X}_n$ be i.i.d. copies of $\boldsymbol{X} \sim H$ with continuous margins $F_1, \dots, F_d$ and corresponding copula C.
- Assume
  - $F_j = F_j(\cdot; \boldsymbol{\theta}_{0,j})$ for some $\boldsymbol{\theta}_{0,j} \in \Theta_j$, $j \in \{1, \dots, d\}$; $(F_j(\cdot; \boldsymbol{\theta}_j)$ continuous $\forall\, \boldsymbol{\theta}_j \in \Theta_j$, $j \in \{1, \dots, d\})$
  - $C = C(\cdot; \boldsymbol{\theta}_{0,C})$ for some $\boldsymbol{\theta}_{0,C} \in \Theta_C$.

  Thus H has the true but unknown parameter vector $\boldsymbol{\theta}_0 = (\boldsymbol{\theta}_{0,C}^\top, \boldsymbol{\theta}_{0,1}^\top, \dots, \boldsymbol{\theta}_{0,d}^\top)^\top$ to be estimated.
- Here, we focus particularly on $\boldsymbol{\theta}_{0,C}$. 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., $\boldsymbol{\theta}_{0,C} = \theta_{0,C}$.
- For $d = 2$, Genest and Rivest (1993) suggested to estimate $\theta_{0,C}$ by solving $\tau(\theta_C) = \hat{\tau}_n$ w.r.t. θ_C, i.e.,

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

  where $\tau(\cdot)$ denotes Kendall's tau as a function in θ and $\hat{\tau}_n$ is the sample version of Kendall's tau (computed via (38) from $\boldsymbol{X}_1, \dots, \boldsymbol{X}_n$ or pseudo-observations $\boldsymbol{U}_1, \dots, \boldsymbol{U}_n$; 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 $\hat{\tau}_{n,j_1 j_2}$ denotes the sample version of Kendall's tau for data pair (j_1, j_2), then $\hat{P}_{n,j_1 j_2}^{\mathsf{IKTE}} = \sin(\frac{\pi}{2}\hat{\tau}_{n,j_1 j_2})$; see Proposition 7.38. For obtaining a proper correlation matrix P (positive semi-definite), see the eigenvalue method (McNeil et al. (2015, Algorithm 7.57)) or Higham (2002) (`Matrix::nearPD()`).

- Extensions to $d > 2$ are used for one-parameter exchangeable copula models; see Berg (2009) or Kojadinovic and Yan (2010). The corresponding *pairwise inversion of Kendall's tau estimator* is

$$\hat{\theta}_{n,C}^{\mathsf{IKTE}} = \tau^{-1}\left(\binom{d}{2}^{-1} \sum_{1 \le j_1 < j_2 \le d} \hat{\tau}_{n,j_1 j_2}\right),$$

where $\hat{\tau}_{n,j_1 j_2}$ is the inversion of Kendall's tau estimator for pair (j_1, j_2).

Remark 7.46

- Clayton copula: $\tau^{-1}(x) = \frac{2x}{1-x}$; Gumbel copula: $\tau^{-1}(x) = \frac{1}{1-x}$.
- Gauss copula: Here we could also use Spearman's rho based on

$$\rho_{\mathsf{S}} \underset{\text{Th.7.36}}{=} \frac{6}{\pi} \arcsin \frac{\rho}{2} \approx \rho.$$

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

- t copula: Use $\hat{P}_n^{\mathsf{IKTE}}$ to estimate P and maximize the likelihood in ν.

7.5.2 Forming a pseudo-sample from the copula

- $\boldsymbol{X}_1, \dots, \boldsymbol{X}_n$ (as good as) never has $\mathrm{U}[0,1]$ margins. For applying the "copula approach" we thus need *pseudo-observations* from C.
- In general, we take $\hat{\boldsymbol{U}}_1, \dots, \hat{\boldsymbol{U}}_n$ with

$$\hat{\boldsymbol{U}}_i = (\hat{U}_{i1}, \dots, \hat{U}_{id}) = (\hat{F}_1(X_{i1}), \dots, \hat{F}_d(X_{id})),$$

  where $\hat{F}_j$ denotes an estimator of F_j; see Lemma 7.9. Note that $\hat{\boldsymbol{U}}_1, \dots, \hat{\boldsymbol{U}}_n$ are typically neither independent (even if $\boldsymbol{X}_1, \dots, \boldsymbol{X}_n$ are) nor perfectly $\mathrm{U}[0,1]$.
- 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

$$\hat{U}_{ij} = \frac{n}{n+1}\hat{F}_{n,j}(X_{ij}) = \frac{R_{ij}}{n+1}, \tag{39}$$

  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 (39) for estimating $\boldsymbol{\theta}_{0,C}$ (to keep the error due to misspecification of the margins small).

3) EVT-based. Bodies are modeled empirically; tails semiparametrically via GPD.

7.5.3 Maximum likelihood estimation

The (classical) maximum likelihood estimator

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

$$h(\boldsymbol{x};\boldsymbol{\theta}_0) = c(F_1(x_1;\boldsymbol{\theta}_{0,1}),\dots,F_d(x_d;\boldsymbol{\theta}_{0,d});\boldsymbol{\theta}_{0,C})\prod_{j=1}^{d} f_j(x_j;\boldsymbol{\theta}_{0,j}).$$

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

$$\ell(\boldsymbol{\theta};\boldsymbol{X}_1,\dots,\boldsymbol{X}_n) = \sum_{i=1}^{n}\ell(\boldsymbol{\theta};\boldsymbol{X}_i)$$

$$= \sum_{i=1}^{n} \ell_C(\boldsymbol{\theta}_C; F_1(X_{i1}; \boldsymbol{\theta}_1), \dots, F_d(X_{id}; \boldsymbol{\theta}_d)) + \sum_{i=1}^{n} \sum_{j=1}^{d} \ell_j(\boldsymbol{\theta}_j; X_{ij}), \tag{40}$$

where

$$\ell_C(\boldsymbol{\theta}_C; u_1, \dots, u_d) = \log c(u_1, \dots, u_d; \boldsymbol{\theta}_C)$$
$$\ell_j(\boldsymbol{\theta}_j; x) = \log f_j(x; \boldsymbol{\theta}_j), \quad j \in \{1, \dots, d\}.$$

- The *maximum likelihood estimator (MLE)* of $\boldsymbol{\theta}_0$ is

$$\hat{\boldsymbol{\theta}}_n^{\mathsf{MLE}} = \operatorname*{argsup}_{\boldsymbol{\theta} \in \Theta} \ell(\boldsymbol{\theta}; \boldsymbol{X}_1, \dots, \boldsymbol{X}_n). \tag{41}$$

  The optimization in (41) is typically done by numerical means. Note that this can be quite demanding, especially in high dimensions.

The inference functions for margins estimator

- Due to the decomposition (40), Joe and Xu (1996) suggested the two-step estimation approach:

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

Step 2: Estimate $\boldsymbol{\theta}_{0,C}$ by

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

The corresponding *inference functions for margins estimator (IFME)* of $\boldsymbol{\theta}_0$ is thus

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

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

$$\hat{\boldsymbol{\theta}}_n^{\text{MLE}} \text{ solves } \left(\frac{\partial}{\partial \boldsymbol{\theta}_C}\ell, \frac{\partial}{\partial \boldsymbol{\theta}_1}\ell, \dots, \frac{\partial}{\partial \boldsymbol{\theta}_d}\ell\right) = \mathbf{0},$$

$$\hat{\boldsymbol{\theta}}_n^{\mathsf{IFME}} \text{ solves } \left(\frac{\partial}{\partial \boldsymbol{\theta}_C}\ell, \frac{\partial}{\partial \boldsymbol{\theta}_1}\ell_1, \dots, \frac{\partial}{\partial \boldsymbol{\theta}_d}\ell_d\right) = \mathbf{0},$$

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

Example 7.47 (A computationally convincing example)

Suppose $X_j \sim \mathrm{N}(\mu_j, \sigma_j^2)$, $j \in \{1, \dots, d\}$, 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 $\boldsymbol{\theta}_{0,C}$ (see MPLE below) in order to avoid misspecifiation 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 as $\hat{\boldsymbol{\theta}}_n^{\mathsf{IFME}}$, but estimates the margins non-parametrically:

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

  Step 2: Estimate $\boldsymbol{\theta}_{0,C}$ by

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

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

Example 7.48 (Fitting the Gauss copula)

- The (copula-related) log-likelihood ℓ_C is
$$\ell_C(P;\hat{\boldsymbol{U}}_1,\dots,\hat{\boldsymbol{U}}_n)=\sum_{i=1}^n \ell_C(P;\hat{\boldsymbol{U}}_i) \underset{\text{Eq. (36)}}{=} \sum_{i=1}^n \log c_P^{\text{Ga}}(\hat{\boldsymbol{U}}_i).$$
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 $\mathcal{O}(d^2)$.
- An approximate solution can be found via $c_P^{\text{Ga}}(\boldsymbol{u}) \underset{\text{Sklar}}{=} \frac{h_P(\Phi^{-1}(u_1),\dots,\Phi^{-1}(u_d))}{\prod_{j=1}^d \varphi(\Phi^{-1}(u_j))}$, where h_P denotes the density of a $\mathrm{N}(\boldsymbol{0},P)$ distribution. Thus
$$\underset{P}{\operatorname{argsup}}\, \ell_C(P;\hat{\boldsymbol{U}}_1,\dots,\hat{\boldsymbol{U}}_n)=\underset{P}{\operatorname{argsup}}\, h_P(\hat{\boldsymbol{Y}}_1,\dots,\hat{\boldsymbol{Y}}_n),\ \ \hat{\boldsymbol{Y}}_i=\Phi^{-1}(\hat{\boldsymbol{U}}_i).$$
Now optimize over all covariance matrices Σ (assuming $\hat{\boldsymbol{Y}}_i$ to be independent!) to obtain the analytic solution $\hat{\Sigma}=\frac{1}{n(-1)}\sum_{i=1}^n \hat{\boldsymbol{Y}}_i\hat{\boldsymbol{Y}}_i^\top$ which is typically close to being a correlation matrix. Thus take $\hat{P}=\mathrm{Cor}[\hat{\Sigma}]$.
- Alternatively, use pairwise inversion of Spearman's rho or Kendall's tau.

Example 7.49 (Fitting the t copula)

- For small d, maximize the likelihood (as for the Gauss copula case) over all correlation matrices (via reparameterization in terms of the Cholesky factor A) 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 $\hat{\nu}$.

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.

A graphical goodness-of-fit approach by Hofert and Mächler (2014); see `demo(gof_graph)` (daily log-returns of SMI, 2011-09-09–2012-03-28).