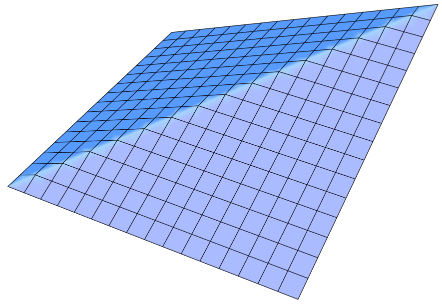


STAT 946: An Introduction to Copula Modeling

Prof. Marius Hofert • Spring 2016 • University of Waterloo
MW 11:30AM–12:50PM in M3 3103

Transcribed, LaTeXed, and edited by Rob Zimmerman

Note: These are personal lecture notes, not official course materials. They may contain errors (which by default you should attribute to me, Rob Zimmerman, rather than the course instructor). The permanent home of these — and all of my other lecture notes — is `https://rob-zimmerman.github.io/coursenotes`.

Contents

Lecture 1 — Monday, May 02, 2016

1. Multivariate Distribution Theory

Preliminaries

The central message of copula modelling is simple: separate marginal behaviour from dependence, and do not let the margins distract from the dependence structure.

Unless otherwise stated, all random variables and random vectors are defined on a common probability space $(\Omega, \mathcal{F}, \mathbb{P})$. When a function is evaluated at $-\infty$ or ∞, its value is understood as the corresponding improper limit, which may itself be infinite.

Let $\boldsymbol{X} = (X_1, \ldots, X_d)\colon \Omega \to \mathbb{R}^d$ be a d-dimensional random vector. Its **distribution function** $H\colon \mathbb{R}^d \to [0,1]$ is

$$H(\boldsymbol{x}) := \mathbb{P}(\boldsymbol{X} \le \boldsymbol{x}), \quad \boldsymbol{x} \in \mathbb{R}^d.$$

For $j \in \{1, \dots, d\}$, the ***j*th marginal distribution function** is

$$F_j(x) := H(\infty, \dots, \infty, x, \infty, \dots, \infty), \quad x \in \mathbb{R},$$

where x occupies the jth coordinate. Similarly, the **survival function** of H is

$$\overline{H}(\boldsymbol{x}) := \mathbb{P}(\boldsymbol{X} > \boldsymbol{x}), \quad \boldsymbol{x} \in \mathbb{R}^d,$$

and its jth marginal survival function is

$$\overline{F}_j(x) := \overline{H}(-\infty, \dots, -\infty, x, -\infty, \dots, -\infty), \quad x \in \mathbb{R}.$$

Remark 1.1 In more than one dimension, $\overline{H}$ is generally not equal to $1 - H$. The bivariate case is illustrated in Figure 1. The lower-left region represents $H(x_1, x_2)$, the upper-right region represents $\overline{H}(x_1, x_2)$, and the complement of the lower-left region represents $1 - H(x_1, x_2)$. Only in the univariate case do we have $\overline{F} = 1 - F$.

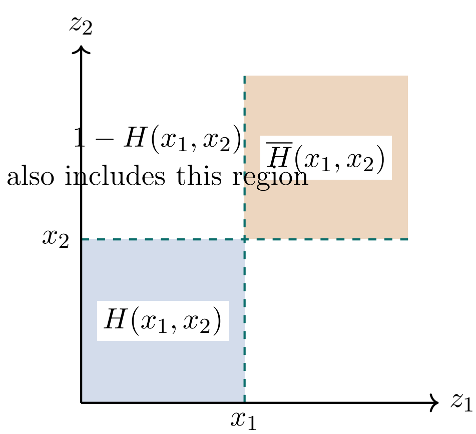


FIGURE 1

A function $T \colon \mathbb{R} \to \mathbb{R}$ is **increasing**, denoted by $T \nearrow$, if $T(x) \le T(y)$ whenever $x \le y$. It is **decreasing**, denoted by $T \searrow$, if $T(x) \ge T(y)$ whenever $x \le y$. Finally, it is **strictly increasing**, denoted by $T \uparrow$, if $T(x) < T(y)$ whenever $x < y$.

Definition 1.2 Let $T \colon \mathbb{R} \to \mathbb{R}$ be increasing. The **generalized inverse** $T^- \colon \mathbb{R} \to \overline{\mathbb{R}}$ is defined by

$$T^-(y) := \inf\{x \in \mathbb{R} : T(x) \ge y\},$$

with the convention $\inf \emptyset := \infty$. If $T\colon \mathbb{R} \to [0,1]$ is a distribution function, then $T^-\colon [0,1] \to \overline{\mathbb{R}}$ is called the **quantile function** of T. Jumps of T correspond to flat portions of T^-, and conversely. If T is continuous and strictly increasing, then

$$T^-(y) = T^{-1}(y) \quad \text{for every } y \in \operatorname{ran}(T),$$

where T^{-1} is the classical inverse of T.

Remark 1.3 The geometry in Figure 2 is worth knowing in one's sleep. In particular, the horizontal segment in the graph of T produces a jump in T^-.

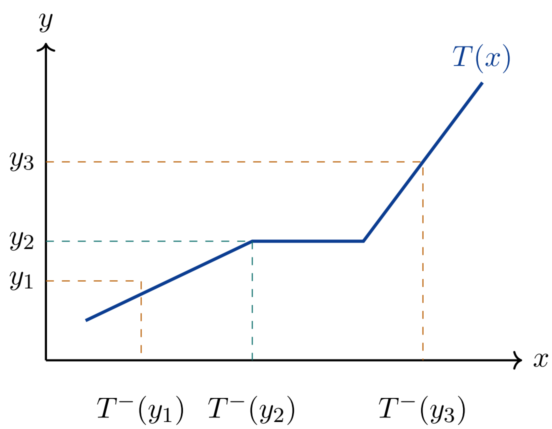


FIGURE 2

The next result collects the main properties of generalized inverses. Their proofs require some care; a detailed treatment is given by Embrechts and Hofert (2013).

Lemma 1.4 Let $T\colon \mathbb{R} \to \mathbb{R}$ be increasing, and let $x, y \in \mathbb{R}$. Then the following statements hold.

1. We have $T^-(y) = -\infty$ if and only if $T(x) \geq y$ for every $x \in \mathbb{R}$. Similarly, $T^-(y) = \infty$ if and only if $T(x) < y$ for every $x \in \mathbb{R}$.
2. If $T^-(y) \in \mathbb{R}$, then T^- is left-continuous at y and has a right-hand limit at y.
3. We have $T^-(T(x)) \leq x$. If T is strictly increasing, then $T^-(T(x)) = x$.

4. If T is right-continuous, then

$$T^{-}(y) \leq x \Longleftrightarrow T(x) \geq y.$$

Moreover, $T(T^{-}(y)) = y$ whenever

$$y \in \operatorname{ran}(T) \cup \{\inf \operatorname{ran}(T), \sup \operatorname{ran}(T)\}.$$

5. If $y < \inf \operatorname{ran}(T)$, then $T(T^{-}(y)) > y$, whereas if $y > \sup \operatorname{ran}(T)$, then $T(T^{-}(y)) < y$. The converses hold when T is right-continuous. In addition, $T(x) < y$ implies $x < T^{-}(y)$.

6. With

$$T^{-}(y-) \coloneqq \lim_{z \nearrow y} T^{-}(z) \qquad \text{and} \qquad T^{-}(y+) \coloneqq \lim_{z \searrow y} T^{-}(z),$$

we have

$$\big(T^{-}(y-), T^{-}(y+)\big) \subseteq \{x \in \mathbb{R} : T(x) = y\} \subseteq \big[T^{-}(y-), T^{-}(y+)\big].$$

7. The function T is continuous if and only if T^{-} is strictly increasing on

$$[\inf \operatorname{ran}(T), \sup \operatorname{ran}(T)].$$

The function T is strictly increasing if and only if T^{-} is continuous on $\operatorname{ran}(T)$.

8. If T_1 and T_2 are right-continuous increasing functions, then

$$(T_2 \circ T_1)^{-} = T_1^{-} \circ T_2^{-}.$$

Lecture 2 — Wednesday, May 04, 2016

For a random variable X, write

$$\operatorname{ran}(X) \coloneqq \{x \in \mathbb{R} : \mathbb{P}(X \in N_x) > 0 \text{ for every open neighbourhood } N_x \text{ of } x\}$$

for its topological support. This set is closed and $X \in \operatorname{ran}(X)$ almost surely. Unlike the distribution function's numerical range, the support may have gaps; the distribution function can therefore be constant between two of its points.

Proposition 2.1 Let F be a distribution function, and suppose that $X \sim F$. Then the following statements hold.

1. If F is continuous, then $F(X) \sim \mathrm{Unif}[0,1]$. This is the **probability transformation**.
2. If $U \sim \mathrm{Unif}[0,1]$, then $F^{-}(U) \sim F$. This is the **quantile transformation**.

Proof. We first prove (2). By Lemma 1.4(4),

$$\mathbb{P}(F^{-}(U) \leq x) = \mathbb{P}(U \leq F(x)) = F(x)$$

for every $x \in \mathbb{R}$. Hence $F^{-}(U) \sim F$.

Now suppose that F is continuous. The intermediate value theorem shows that every $u \in (0,1)$ belongs to the range of F, and Lemma 1.4(4) therefore gives

$$F(F^{-}(u)) = u.$$

Consequently, if $U \sim \mathrm{Unif}[0,1]$, then $F(F^{-}(U)) = U$ almost surely. Since the first part of the proof gives $F^{-}(U) \overset{\mathrm{d}}{=} X$, we conclude that $F(X) \overset{\mathrm{d}}{=} U$. This proves (1) without requiring F to be strictly increasing; a continuous distribution function may be constant across gaps in its support. $\square$

Proposition 2.1(1) can be used for estimation and goodness-of-fit testing. Suppose that $X_1, \ldots, X_n$ are independent and identically distributed with an unknown continuous distribution function F. To test $H_0\colon F = F_0$, we can instead test whether $F_0(X_1), \ldots, F_0(X_n)$ are independent and uniformly distributed on $[0,1]$, for example by using a Kolmogorov–Smirnov or Anderson–Darling test.

Proposition 2.1(2) can be used to sample nonuniform random variables. If F^{-} has a convenient form, the Mersenne Twister may be used to generate independent pseudorandom variables $U_i \sim \mathrm{Unif}[0,1]$; then the variables $F^{-}(U_i)$ are independent with distribution function F.

Both tasks become considerably harder in more than one dimension. The random variable $F(X)$ in Proposition 2.1(1) is the probability integral transform. It is guaranteed to be uniform when F is continuous, and the next definition extends it to discontinuous distributions.

Definition 2.2 Let $X \sim F$. The **modified distribution function** is defined by

$$F(x,\lambda) \coloneqq \mathbb{P}(X < x) + \lambda\mathbb{P}(X = x) = F(x-) + \lambda\big(F(x) - F(x-)\big)$$

for $x \in \mathbb{R}$ and $\lambda \in [0,1]$. The **generalized distributional transform** of X is $F(X,\Delta)$, where $\Delta \sim \mathrm{Unif}[0,1]$ is independent of X. Informally, $F(X,\Delta)$ uniformly fills the gaps in the range of F: if F jumps at x, then $F(x,\Delta)$ is uniformly distributed on $[F(x-), F(x)]$.

Proposition 2.3 Let $X \sim F$, and let $\Delta \sim \mathrm{Unif}[0,1]$ be independent of X. Then the following statements hold.

1. If F is continuous, then $F(X,\Delta) = F(X)$.
2. We have $F(X,\Delta) \sim \mathrm{Unif}[0,1]$.
3. We have $F^{-}(F(X,\Delta)) = X$ almost surely.

Proof. For (1), continuity of F gives

$$F(x,\lambda) = F(x-) + \lambda\big(F(x) - F(x-)\big) = F(x).$$

For (2), fix $y \in (0,1)$ and write $x_y \coloneqq F^{-}(y)$. The event $\{F(X,\Delta) \le y\}$ is the disjoint union

$$\{X < x_y\} \,\dot{\cup}\, \big\{X = x_y,\, \mathbb{P}(X < x_y) + \Delta\mathbb{P}(X = x_y) \le y\big\}.$$

If $\mathbb{P}(X = x_y) = 0$, the defining properties of x_y imply that

$$\mathbb{P}(X < x_y) = \mathbb{P}(X \le x_y) = y.$$

If $\mathbb{P}(X = x_y) > 0$, independence and the uniform distribution of Δ give

$$\begin{aligned}\mathbb{P}(F(X,\Delta) \le y) &= \mathbb{P}(X < x_y) \\ &\quad + \mathbb{P}(X = x_y)\mathbb{P}\left(\Delta \le \frac{y - \mathbb{P}(X < x_y)}{\mathbb{P}(X = x_y)}\right) \\ &= \mathbb{P}(X < x_y) + y - \mathbb{P}(X < x_y) \\ &= y.\end{aligned}$$

Thus $F(X,\Delta)$ is uniformly distributed on $[0,1]$.

For (3), first suppose that x is an atom of F. Then

$$F^{-}(y) = x \quad \text{for every } y \in (F(x-), F(x)].$$

Because $\Delta > 0$ almost surely, the asserted identity therefore holds on the event that X is an atom. At any non-atom x, $F(x, \Delta) = F(x)$. The identity $F^{-}(F(x)) = x$ can fail at such an x only when x is an endpoint of a nonempty interval on which F is constant. There are at most countably many disjoint maximal intervals of constancy, and their non-atomic endpoints have probability zero. Hence

$$F^{-}(F(X, \Delta)) = X$$

almost surely. □

Lecture 3 — Monday, May 09, 2016

2. Copulas and Sklar's Theorem

Copulas

Definition 3.1 Let $H\colon \mathbb{R}^d \to [0,1]$, and let $\boldsymbol{x}, \boldsymbol{a}, \boldsymbol{b} \in \mathbb{R}^d$ with $\boldsymbol{a} \leq \boldsymbol{b}$.

1. The function H is **grounded** if

   $$H(x_1, \ldots, x_{j-1}, -\infty, x_{j+1}, \ldots, x_d) = 0$$

   for every $j \in \{1, \ldots, d\}$.
2. The jth **margin** of H is

   $$F_j(x) \coloneqq H(\infty, \ldots, \infty, x, \infty, \ldots, \infty),$$

   where x occupies the jth coordinate.

3. Define the increment in the jth coordinate by

$$\Delta_{(a_j,b_j]} H(\boldsymbol{x}) \coloneqq H(x_1,\ldots,b_j,\ldots,x_d) - H(x_1,\ldots,a_j,\ldots,x_d).$$

The ***H*-volume** of $(\boldsymbol{a},\boldsymbol{b}]$ is

$$\begin{aligned}\Delta_{(\boldsymbol{a},\boldsymbol{b}]} H &\coloneqq \Delta_{(a_1,b_1]} \cdots \Delta_{(a_d,b_d]} H \\ &= \sum_{i_1=1}^{2} \cdots \sum_{i_d=1}^{2} (-1)^{\sum_{j=1}^{d} i_j} H(x_{1,i_1},\ldots,x_{d,i_d}),\end{aligned}$$

where

$$x_{j,i_j} = \begin{cases} a_j, & i_j = 1, \\ b_j, & i_j = 2. \end{cases}$$

4. The function H is ***d*-increasing** if $\Delta_{(\boldsymbol{a},\boldsymbol{b}]} H \geq 0$ for all $\boldsymbol{a},\boldsymbol{b} \in \mathbb{R}^d$ satisfying $\boldsymbol{a} \leq \boldsymbol{b}$.

Remark 3.2 Being d-increasing is much stronger than being increasing in each coordinate. When $d = 2$, the H-volume in Figure 3 is

$$\Delta_{(\boldsymbol{a},\boldsymbol{b}]} H = H(b_1,b_2) - H(a_1,b_2) - H(b_1,a_2) + H(a_1,a_2).$$

Thus a 2-increasing function assigns nonnegative mass to every axis-aligned rectangle.

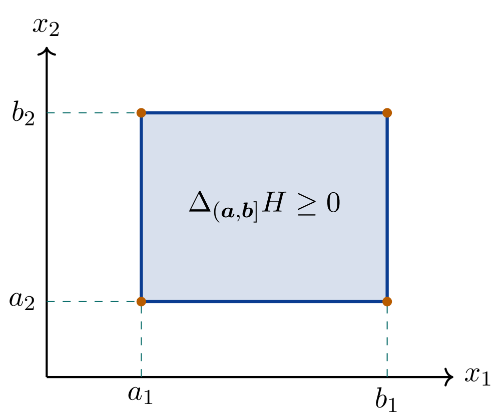


Figure 3

Proposition 3.3 Let H be grounded and d-increasing. Then the following statements hold.

1. Let $k \in \{1, \dots, d\}$ and $\{j_1, \dots, j_k\} \subseteq \{1, \dots, d\}$. Then
$$\Delta_{(a_{j_1}, b_{j_1}]} \cdots \Delta_{(a_{j_k}, b_{j_k}]} H(\boldsymbol{x}) \geq 0$$
for every $\boldsymbol{x} \in \mathbb{R}^d$.
2. $$|H(\boldsymbol{b}) - H(\boldsymbol{a})| \leq \sum_{j=1}^{d} |F_j(b_j) - F_j(a_j)|$$
for all $\boldsymbol{a}, \boldsymbol{b} \in \mathbb{R}^d$.

Proof. For (1), extend $a_{j_1}, \dots, a_{j_k}$ to a vector $\boldsymbol{a}$ by setting $a_j = a_{j_\ell}$ when $j = j_\ell$ for some $\ell \in \{1, \dots, k\}$, and setting $a_j = -\infty$ otherwise. Similarly, extend $b_{j_1}, \dots, b_{j_k}$ to a vector $\boldsymbol{b}$ by setting $b_j = b_{j_\ell}$ when $j = j_\ell$, and setting $b_j = x_j$ otherwise. Since H is grounded, all terms containing a coordinate equal to $-\infty$ vanish. Hence

$$\Delta_{(a_{j_1}, b_{j_1}]} \cdots \Delta_{(a_{j_k}, b_{j_k}]} H(\boldsymbol{x}) = \Delta_{(\boldsymbol{a}, \boldsymbol{b}]} H \geq 0,$$

where the inequality follows from the d-increasingness of H.

For (2), first use a telescoping sum and the triangle inequality to obtain

$$|H(\boldsymbol{b}) - H(\boldsymbol{a})| \leq \sum_{j=1}^{d} \big| H(b_1, \dots, b_j, a_{j+1}, \dots, a_d) \\ - H(b_1, \dots, b_{j-1}, a_j, \dots, a_d) \big|.$$

Assume first that $a_j \leq b_j$. Part (1), with $k = 1$, shows that the difference inside the absolute value is nonnegative. Repeatedly applying the same part with $k = 2$ gives

$$\begin{aligned}
&\big| H(b_1, \dots, b_j, \dots, a_d) - H(b_1, \dots, a_j, \dots, a_d) \big| \\
&\quad = H(b_1, \dots, b_j, \dots, a_d) - H(b_1, \dots, a_j, \dots, a_d) \\
&\quad \leq H(\infty, \dots, b_j, \dots, \infty) - H(\infty, \dots, a_j, \dots, \infty) \\
&\quad = F_j(b_j) - F_j(a_j) = |F_j(b_j) - F_j(a_j)|.
\end{aligned}$$

The same inequality follows after interchanging a_j and b_j when $a_j > b_j$. Summing over j proves the claim. □

Whenever the indicated mixed partial derivative exists, write

$$D_{j_k,\ldots,j_1} H(\boldsymbol{x}) \coloneqq \left. \frac{\partial^k H(\boldsymbol{z})}{\partial z_{j_k} \cdots \partial z_{j_1}} \right|_{\boldsymbol{z}=\boldsymbol{x}}.$$

Proposition 3.4 Let $H\colon \mathbb{R}^d \to [0,1]$ be grounded and d-increasing. Fix $j \in \{1,\ldots,d\}$ and all coordinates other than x_j. Then $D_j H(\boldsymbol{x})$ exists for almost every $x_j \in \mathbb{R}$ with respect to Lebesgue measure, and

$$D_j H(\boldsymbol{x}) \geq 0,$$

wherever it exists. More generally, fix finitely many endpoint configurations in coordinates other than j. For almost every x_j, all the corresponding derivatives exist simultaneously, and applying any increment formed from those endpoints gives a nonnegative result. In particular, for $k \neq j$ and $a_k \leq b_k$,

$$D_j H(\boldsymbol{x}_{-k}, b_k) \geq D_j H(\boldsymbol{x}_{-k}, a_k)$$

whenever both derivatives exist.

Proof. Proposition 3.3(1) implies that H is increasing in x_j. A monotone real-valued function is differentiable almost everywhere, so $D_j H(\boldsymbol{x})$ exists for almost every x_j with respect to Lebesgue measure.

Fix increments in any collection of coordinates other than j, and denote their product by Δ_{-j}. Set

$$G(x_j) \coloneqq \Delta_{-j} H(\boldsymbol{x}).$$

The d-increasingness of H implies that G is increasing in x_j. Consequently, wherever the derivative exists,

$$\Delta_{-j} D_j H(\boldsymbol{x}) = D_j G(x_j) \geq 0.$$

Only finitely many coordinate configurations occur in this calculation, so their full-measure differentiability sets have a full-measure intersection. The first assertion also follows directly because H is increasing in x_j; using a single increment in coordinate k gives the final inequality. □

These multivariate distribution properties lead directly to copulas.

Definition 3.5 A d-dimensional **copula** is a d-dimensional distribution function with standard uniform margins.

Proposition 3.6 A function $C\colon [0,1]^d \to [0,1]$ is a d-dimensional copula if and only if the following conditions hold.

1. The function C is grounded: $C(\boldsymbol{u}) = 0$ whenever $u_j = 0$ for some $j \in \{1, \dots, d\}$.
2. The function C has standard uniform margins: $C(1, \dots, 1, u_j, 1, \dots, 1) = u_j$ for every $u_j \in [0,1]$ and every $j \in \{1, \dots, d\}$.
3. The function C is d-increasing: $\Delta_{(\boldsymbol{a},\boldsymbol{b}]} C \geq 0$ for all $\boldsymbol{a}, \boldsymbol{b} \in [0,1]^d$ satisfying $\boldsymbol{a} \leq \boldsymbol{b}$.

Proof. *Forward implication.* Let $C\colon [0,1]^d \to [0,1]$ be a copula, and let $\boldsymbol{U} \sim C$. For (1),

$$0 \leq C(u_1, \dots, u_{j-1}, 0, u_{j+1}, \dots, u_d) \leq \mathbb{P}(U_j = 0) = 0.$$

For (2),

$$C(1, \dots, 1, u_j, 1, \dots, 1) = \mathbb{P}(U_j \leq u_j) = u_j$$

for every $j \in \{1, \dots, d\}$. Finally, for (3),

$$\Delta_{(\boldsymbol{a},\boldsymbol{b}]} C = \mathbb{P}(\boldsymbol{U} \in (\boldsymbol{a}, \boldsymbol{b}]) \geq 0$$

whenever $\boldsymbol{a}, \boldsymbol{b} \in [0,1]^d$ and $\boldsymbol{a} \leq \boldsymbol{b}$.

Reverse implication. Condition (2) ensures the correct margins once C is known to be a distribution function, so it remains to prove that C is a distribution function. Write

$$x \wedge y := \min\{x, y\} \qquad \text{and} \qquad x \vee y := \max\{x, y\}.$$

Extend C to $H\colon \mathbb{R}^d \to [0,1]$ by setting

$$H(\boldsymbol{x}) := C\big((0 \vee x_1) \wedge 1, \dots, (0 \vee x_d) \wedge 1\big).$$

Condition (1) gives

$$H(x_1, \dots, x_{j-1}, -\infty, x_{j+1}, \dots, x_d) = 0,$$

whereas Condition (2) gives

$$H(\infty, \ldots, \infty) = C(1, \ldots, 1) = 1.$$

For $\boldsymbol{a}, \boldsymbol{b} \in \mathbb{R}^d$ with $\boldsymbol{a} \leq \boldsymbol{b}$, define

$$\widetilde{a}_j = (0 \vee a_j) \wedge 1 \qquad \text{and} \qquad \widetilde{b}_j = (0 \vee b_j) \wedge 1.$$

Then Condition (3) implies that

$$\Delta_{(\boldsymbol{a},\boldsymbol{b}]} H = \Delta_{(\widetilde{\boldsymbol{a}},\widetilde{\boldsymbol{b}}]} C \geq 0.$$

Thus H is grounded and d-increasing. By Proposition 3.3(2),

$$\begin{aligned}
|H(\boldsymbol{x} + \boldsymbol{h}) - H(\boldsymbol{x})| &\leq \sum_{j=1}^{d} |F_j(x_j + h_j) - F_j(x_j)| \\
&= \sum_{j=1}^{d} |(0 \vee (x_j + h_j)) \wedge 1 - (0 \vee x_j) \wedge 1| \\
&\leq \sum_{j=1}^{d} |h_j| \longrightarrow 0 \quad \text{as } \boldsymbol{h} \to \boldsymbol{0}.
\end{aligned}$$

Hence H is continuous from above. The extension theorem for distribution functions then gives a unique probability measure $\mathbb{P}_C$ on $(\mathbb{R}^d, \mathcal{B}_{\mathbb{R}^d})$ satisfying

$$\mathbb{P}_C((\boldsymbol{a}, \boldsymbol{b}]) = \Delta_{(\boldsymbol{a},\boldsymbol{b}]} H.$$

This is the extension theorem presented in Billingsley (1995, pp. 179 and 260). Therefore, C is the restriction to $[0, 1]^d$ of the distribution function associated with $\mathbb{P}_C$, since

$$C(\boldsymbol{u}) = \mathbb{P}_C((\boldsymbol{0}, \boldsymbol{u}]).$$

□

Lecture 4 — Wednesday, May 11, 2016

Remark 4.1

1. Proposition 3.3(2) gives
$$|C(\boldsymbol{b}) - C(\boldsymbol{a})| \leq \sum_{j=1}^{d} |b_j - a_j|$$
for all $\boldsymbol{a}, \boldsymbol{b} \in [0,1]^d$. Thus every copula is Lipschitz continuous with constant 1 relative to the ℓ_1-norm, and the family of all copulas is uniformly equicontinuous.

2. By Proposition 3.4, $D_j C(\boldsymbol{u})$ exists for almost every u_j with respect to Lebesgue measure, and
$$0 \leq D_j C(\boldsymbol{u}) \leq 1.$$
Moreover, the map $u_j \mapsto D_j C(\boldsymbol{u})$ is Lebesgue integrable.

3. Let C be a copula, and let $\boldsymbol{U} \sim C$. For each $j \in \{1, \ldots, d\}$, continuity from above gives
$$\begin{aligned} 0 \leq \mathbb{P}(\boldsymbol{U} = \boldsymbol{u}) &\leq \mathbb{P}(U_j = u_j) \\ &= \mathbb{P}\left(U_j \in \bigcap_{n\in\mathbb{N}} \left(u_j - \frac{1}{n}, u_j\right]\right) \\ &= \lim_{n\to\infty} \mathbb{P}\left(U_j \in \left(u_j - \frac{1}{n}, u_j\right]\right) \\ &= 0. \end{aligned}$$
Therefore, the probability measure $\mathbb{P}_C$ associated with C has no point masses. Nevertheless, $\mathbb{P}_C$ may have a **singular component**: a nonzero component of the measure supported on a subset of $[0,1]^d$ with Lebesgue measure zero. This happens for the copulas W, M, and the Marshall–Olkin family. If all probability mass belongs to the singular component, then C is itself called **singular**.

4. If C_1 and C_2 are copulas and $\lambda \in [0,1]$, then
$$C(\boldsymbol{u}) := \lambda C_1(\boldsymbol{u}) + (1-\lambda) C_2(\boldsymbol{u})$$

is also a copula. Thus the set of copulas is convex.

Analytic proof. For $j \in \{1, \dots, d\}$, let

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

Then

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

so C is grounded. Similarly, with $\boldsymbol{u}_j^1 \coloneqq (1, \dots, 1, u_j, 1, \dots, 1)$,

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

Finally,

$$\Delta_{(\boldsymbol{a},\boldsymbol{b}]} C = \lambda \Delta_{(\boldsymbol{a},\boldsymbol{b}]} C_1 + (1-\lambda) \Delta_{(\boldsymbol{a},\boldsymbol{b}]} C_2 \ge 0.$$

By Proposition 3.6, C is a copula. □

A probabilistic proof follows from a stochastic representation. Let $\boldsymbol{U}_1 \sim C_1$, $\boldsymbol{U}_2 \sim C_2$, and $X \sim \mathrm{Bin}(1, \lambda)$ be independent, so that $\mathbb{P}(X = 1) = \lambda$. Define

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

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

Probabilistic proof.

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

□

More generally, if $(C_{\boldsymbol{\theta}})_{\boldsymbol{\theta} \in \mathbb{R}^p}$ is a family of copulas, $\boldsymbol{\theta} \mapsto C_{\boldsymbol{\theta}}(\boldsymbol{u})$ is measurable for every $\boldsymbol{u}$, and F is a mixing distribution, then the Stieltjes integral

$$C(\boldsymbol{u}) = \int_{\mathbb{R}^p} C_{\boldsymbol{\theta}}(\boldsymbol{u}) \, \mathrm{d}F(\boldsymbol{\theta})$$

defines a copula. The verification uses the same argument and the linearity of the integral.

Theorem 4.2 (Fréchet–Hoeffding bounds) Let $d \geq 2$. Define

$$W(\boldsymbol{u}) \coloneqq \max\left\{\sum_{j=1}^{d} u_j - d + 1, 0\right\}$$

and

$$M(\boldsymbol{u}) \coloneqq \min_{1 \leq j \leq d}\{u_j\}.$$

1. Every d-dimensional copula C satisfies

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

2. The function W is a copula if and only if $d = 2$.
3. The function M is a copula for every $d \geq 2$.

Proof. For (1), Proposition 3.3(2) gives

$$1 - C(\boldsymbol{u}) = C(\mathbf{1}) - C(\boldsymbol{u}) \leq \sum_{j=1}^{d}(1 - u_j) = d - \sum_{j=1}^{d} u_j.$$

Together with $C(\boldsymbol{u}) \geq 0$, this proves the lower bound. For the upper bound, monotonicity in each coordinate gives

$$C(\boldsymbol{u}) \leq C(1, \ldots, 1, u_j, 1, \ldots, 1) = u_j$$

for every $j \in \{1, \ldots, d\}$. Hence $C(\boldsymbol{u}) \leq M(\boldsymbol{u})$.

For (2), first prove one direction.

Reverse implication. Suppose that $d = 2$, and let $U \sim \text{Unif}[0,1]$. Then

$$\mathbb{P}(U \leq u_1, 1 - U \leq u_2) = \mathbb{P}(1 - u_2 \leq U \leq u_1) = \max\{u_1 + u_2 - 1, 0\}.$$

Thus $(U, 1-U)$ has distribution function W, so W is a copula. $\square$

Lecture 5 — Monday, May 16, 2016

Proof of Theorem 4.2, continued. *Forward implication.* Suppose that W is a copula. If $d \geq 3$, then its volume on $((1/2)\mathbf{1}, \mathbf{1}]$ is

$$\begin{aligned}
\Delta_{((1/2)\mathbf{1},\mathbf{1}]} W &= \sum_{\boldsymbol{i} \in \{0,1\}^d} (-1)^{d - \sum_{j=1}^d i_j} W\left(\left(\frac{1}{2}\right)^{1-i_1}, \dots, \left(\frac{1}{2}\right)^{1-i_d}\right) \\
&= \sum_{\boldsymbol{i} \in \{0,1\}^d} (-1)^{d - \sum_{j=1}^d i_j} \max\left\{\sum_{j=1}^d \left(\frac{1}{2}\right)^{1-i_j} - d + 1, 0\right\} \\
&= \max\{d - d + 1, 0\} - d \max\left\{\frac{1}{2} + d - 1 - d + 1, 0\right\} \\
&\quad + \binom{d}{2} \max\left\{\frac{1}{2} + \frac{1}{2} + d - 2 - d + 1, 0\right\} - \dots \\
&\quad + (-1)^d \max\left\{\frac{d}{2} - d + 1, 0\right\} \\
&= 1 - \frac{d}{2} < 0.
\end{aligned}$$

This contradicts d-increasingness. Therefore, because $d \geq 2$ throughout, W can be a copula only when $d = 2$. This completes the proof of (2).

For (3), let $U \sim \mathrm{Unif}[0,1]$ and set $\boldsymbol{U} = (U, \dots, U)$. Then

$$\mathbb{P}(\boldsymbol{U} \leq \boldsymbol{u}) = \mathbb{P}\left(U \leq \min_{1 \leq j \leq d} u_j\right) = \min_{1 \leq j \leq d} u_j = M(\boldsymbol{u}).$$

Thus M is a copula in every dimension $d \geq 2$. □

Remark 5.1

1. For $d \geq 3$ and each $\boldsymbol{u} \in [0,1]^d$, we can construct a copula C satisfying $C(\boldsymbol{u}) = W(\boldsymbol{u})$. Thus the lower bound W is pointwise sharp even when it is not itself a copula.
2. The functions W and M are called the lower and upper Fréchet–Hoeffding bounds, respectively.

3. In two dimensions, W distributes its mass uniformly on

$$\{(u_1, u_2) \in [0,1]^2 : u_2 = 1 - u_1\},$$

whereas M distributes its mass uniformly on

$$\{(u_1, u_2) \in [0,1]^2 : u_2 = u_1\}.$$

Thus both are singular copulas and do not possess densities with respect to two-dimensional Lebesgue measure. By contrast, the independence copula $\Pi(u_1, u_2) = u_1 u_2$ has the constant density $c_\Pi = 1$. The three measure geometries are compared in Figure 4.

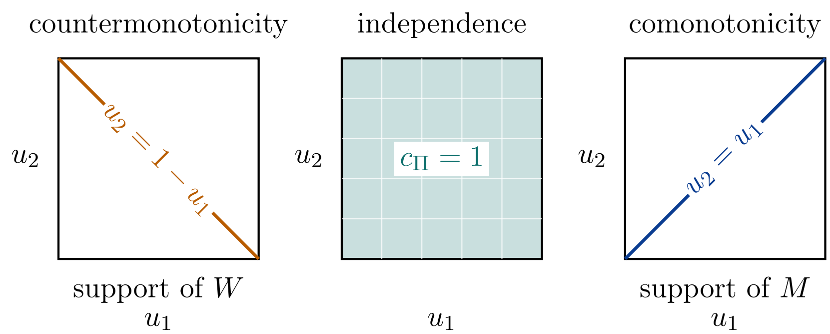


Figure 4

The copula functions themselves are shown in Figure 5. The countermonotone copula W is zero on the lower triangle and rises linearly on the upper triangle, the independence copula Π forms the bilinear saddle $u_1 u_2$, and the comonotone copula M consists of the two planar faces determined by $\min\{u_1, u_2\}$.

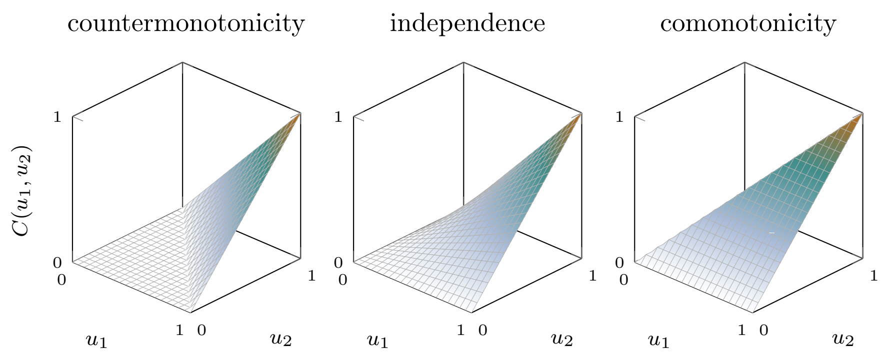


Figure 5

A related geometric calculation in the region $x_1 + x_2 \geq 1$ is

$$\begin{aligned} W_{\text{geom}}(x_1, x_2) &\coloneqq \frac{\sqrt{(x_1 - (1 - x_2))^2 + (x_2 - (1 - x_1))^2}}{\sqrt{2}} \\ &= \frac{\sqrt{2(x_1 + x_2 - 1)^2}}{\sqrt{2}} \\ &= x_1 + x_2 - 1. \end{aligned}$$

This is precisely the positive affine branch of $W(x_1, x_2)$.

Theorem 5.2 (Sklar's theorem)

1. If H is a distribution function with margins $F_1, \dots, F_d$, then there is a copula C such that
$$H(\boldsymbol{x}) = C(F_1(x_1), \dots, F_d(x_d)) \quad \text{for every } \boldsymbol{x} \in \mathbb{R}^d. \tag{5.1}$$
The copula C is uniquely determined on $\prod_{j=1}^{d} \operatorname{ran}(F_j)$, where
$$C(\boldsymbol{u}) = H(F_1^{-}(u_1), \dots, F_d^{-}(u_d)).$$
2. Conversely, if C is any copula and $F_1, \dots, F_d$ are univariate distribution functions, then the function H defined by (5.1) is a distribution function with margins $F_1, \dots, F_d$.

Proof. For (1), let $\boldsymbol{X} \sim H$, and independently let $\boldsymbol{A}$ be uniformly distributed on $[0, 1]^d$. Define $\boldsymbol{U} = (U_1, \dots, U_d)$ by
$$U_j \coloneqq F_j(X_j, A_j), \quad j \in \{1, \dots, d\}.$$
Let C^* be the distribution function of $\boldsymbol{U}$. Proposition 2.3(2) shows that each U_j is uniformly distributed, so C^* is a copula. Proposition 2.3(3) gives $X_j = F_j^{-}(U_j)$ almost surely. Therefore,

$$\begin{aligned} H(\boldsymbol{x}) &= \mathbb{P}\left(\bigcap_{j=1}^{d} \{X_j \leq x_j\}\right) \\ &= \mathbb{P}\left(\bigcap_{j=1}^{d} \{F_j^{-}(U_j) \leq x_j\}\right) \end{aligned}$$

$$
\begin{aligned}
&= \mathbb{P}\left(\bigcap_{j=1}^{d}\{U_j \le F_j(x_j)\}\right) \\
&= C^*(F_1(x_1), \ldots, F_d(x_d)).
\end{aligned}
$$

Hence C^* satisfies (5.1). For $u_j \in \operatorname{ran}(F_j)$, Lemma 1.4(4) gives $F_j(F_j^-(u_j)) = u_j$. It follows that

$$
C^*(\boldsymbol{u}) = C^*\big(F_1(F_1^-(u_1)), \ldots, F_d(F_d^-(u_d))\big) = H(F_1^-(u_1), \ldots, F_d^-(u_d)),
$$

which also proves uniqueness on $\prod_{j=1}^{d} \operatorname{ran}(F_j)$.

For (2), let $\boldsymbol{U} \sim C$ and define

$$
\boldsymbol{X} \coloneqq (F_1^-(U_1), \ldots, F_d^-(U_d)).
$$

Then

$$
\begin{aligned}
\mathbb{P}(\boldsymbol{X} \le \boldsymbol{x}) &= \mathbb{P}\left(\bigcap_{j=1}^{d}\{F_j^-(U_j) \le x_j\}\right) \\
&= \mathbb{P}\left(\bigcap_{j=1}^{d}\{U_j \le F_j(x_j)\}\right) \\
&= C(F_1(x_1), \ldots, F_d(x_d)) \\
&= H(\boldsymbol{x}).
\end{aligned}
$$

Thus H is the distribution function of $\boldsymbol{X}$, and Proposition 2.1(2) shows that its margins are $F_1, \ldots, F_d$. □

Remark 5.3 Theorem 5.2(1) decomposes every multivariate distribution function into its margins and a copula. Consequently, the dependence structure may be studied independently of the margins, which is useful for parameter estimation and goodness-of-fit testing. Theorem 5.2(2) constructs multivariate distribution functions from chosen margins and a copula, which is useful for model building and stress testing.

Lecture 6 — Wednesday, May 18, 2016

Remark 6.1

1. The copula C^*, read as "C Maltese," is also called the **standard extension copula**. The classical proof of Theorem 5.2 constructs this copula analytically by multilinear interpolation. The probabilistic construction used here is due to Rüschendorf (2009).
2. The copula in (5.1) need not be unique when the margins are discontinuous. For example, suppose that (X_1, X_2) has the bivariate Bernoulli distribution

$$\mathbb{P}(X_1 = h, X_2 = \ell) = \frac{1}{4}, \quad h, \ell \in \{0, 1\}.$$

Each margin assigns probability 1/2 to both 0 and 1, so

$$\operatorname{ran}(F_j) = \{0, 1/2, 1\}, \quad j \in \{1, 2\}.$$

Consequently, every copula C satisfying $C(1/2, 1/2) = 1/4$ gives the same joint distribution through (5.1). One choice is the independence copula

$$C_1(u_1, u_2) = \Pi(u_1, u_2) \coloneqq u_1 u_2.$$

Another is the diagonal copula

$$C_2(u_1, u_2) = \min\left\{u_1, u_2, \frac{\delta(u_1) + \delta(u_2)}{2}\right\}, \quad \delta(u) = u^2.$$

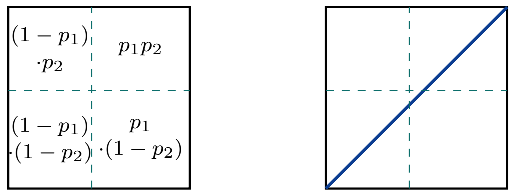
These copulas are distinct because

$$C_1\left(\frac{1}{2}, \frac{1}{3}\right) = \frac{1}{6} \neq \frac{13}{72} = C_2\left(\frac{1}{2}, \frac{1}{3}\right).$$

3. The construction in the proof of Theorem 5.2 also reveals this nonuniqueness. The auxiliary vector $\boldsymbol{A}$ may have either the independence copula Π or the upper Fréchet–Hoeffding copula M, while still producing the same factorization. For a bivariate Bernoulli vector $\boldsymbol{X} \sim H$, write

$$p_j \coloneqq \mathbb{P}(X_j = 1) = 1 - \mathbb{P}(X_j = 0), \quad j \in \{1, 2\}.$$

The two possible mass allocations are illustrated in Figure 6.



$\boldsymbol{A} \sim \Pi$ $\quad\quad$ $\boldsymbol{A} \sim M$

Figure 6

4. If $\boldsymbol{X} \sim H$ has margins $F_1, \ldots, F_d$, then $\boldsymbol{X}$, or equivalently H, is said to have copula C when (5.1) holds.

Lemma 6.2 Suppose that $X_j \sim F_j$ and that each F_j is continuous. Then $\boldsymbol{X}$ has copula C if and only if

$$\boldsymbol{U} := (F_1(X_1), \ldots, F_d(X_d)) \sim C.$$

Proof. *Forward implication.* Let H be the distribution function of $\boldsymbol{X}$, and suppose that $\boldsymbol{X}$ has copula C. For $\boldsymbol{u} \in [0,1]^d$, continuity of the margins and Lemma 1.4 give

$$\begin{aligned}
\mathbb{P}(\boldsymbol{U} \le \boldsymbol{u}) &= \mathbb{P}\left(\bigcap_{j=1}^{d} \{X_j \le F_j^{-}(u_j)\}\right) \\
&= H(F_1^{-}(u_1), \ldots, F_d^{-}(u_d)) \\
&= C(\boldsymbol{u}).
\end{aligned}$$

Here the first event equality is understood up to a null event when a margin is constant on an interval, and continuity gives $F_j(F_j^{-}(u_j)) = u_j$ for $u_j \in (0,1)$; the endpoints follow by continuity. Thus $\boldsymbol{U} \sim C$.

Reverse implication. Suppose that $\boldsymbol{U} \sim C$. Continuity of each margin gives

$$\begin{aligned}
H(\boldsymbol{x}) &= \mathbb{P}\left(\bigcap_{j=1}^{d} \{X_j \le x_j\}\right) \\
&= \mathbb{P}\left(\bigcap_{j=1}^{d} \{F_j(X_j) \le F_j(x_j)\}\right)
\end{aligned}$$

$$= C(F_1(x_1), \ldots, F_d(x_d)).$$

The first event equality again holds up to a null event: points beyond x_j on which F_j remains equal to $F_j(x_j)$ carry no marginal probability. By Theorem 5.2, $\boldsymbol{X}$ has copula C. $\square$

Theorem 6.3 Let $\boldsymbol{X} \sim H$ have continuous margins $F_1, \ldots, F_d$ and copula C. Then the following statements hold.

1. If T_j is measurable and strictly increasing on $\operatorname{ran}(X_j)$ for each $j \in \{1, \ldots, d\}$, then $(T_1(X_1), \ldots, T_d(X_d))$ also has copula C. Thus copulas are invariant under strictly increasing transformations of the underlying random variables.
2. The random variables $X_1, \ldots, X_d$ are independent if and only if

$$C(\boldsymbol{u}) = \Pi(\boldsymbol{u}) := \prod_{j=1}^{d} u_j$$

for every $\boldsymbol{u} \in [0,1]^d$.

3. When $d = 2$, we have $C = W$ if and only if

$$X_2 = F_2^{-} \circ (1 - F_1)(X_1)$$

almost surely. In this case, X_1 and X_2 are **countermonotone**.

4. We have $C = M$ if and only if

$$X_j = F_j^{-} \circ F_1(X_1) \quad \text{almost surely for every } j \in \{2, \ldots, d\}.$$

In this case, $X_1, \ldots, X_d$ are **comonotone**.

Proof. For (1), let G_j be the distribution function of $T_j(X_j)$. The transformed variable has no atoms: strict increase of T_j on $\operatorname{ran}(X_j)$ makes the inverse image of any point contain at most one point of that range, and X_j has no atoms. Thus G_j is continuous. Moreover, for every $x \in \operatorname{ran}(X_j)$, strict increase on the range gives

$$G_j(T_j(x)) = \mathbb{P}(T_j(X_j) \le T_j(x)) = \mathbb{P}(X_j \le x) = F_j(x),$$

where $X_j \in \operatorname{ran}(X_j)$ almost surely. Hence

$$G_j(T_j(X_j)) = F_j(X_j) \quad \text{almost surely.}$$

The marginal probability transforms of the original and transformed vectors therefore agree almost surely. By Lemma 6.2, the transformed vector has copula C.

For (2), Theorem 5.2 gives

$$H(\boldsymbol{x}) = \prod_{j=1}^{d} F_j(x_j) \Longleftrightarrow C(\boldsymbol{u}) = \prod_{j=1}^{d} u_j,$$

where the reverse direction follows by substituting $x_j = F_j^{-}(u_j)$.

For (3), begin with one direction.

Forward implication. Suppose that

$$X_2 = F_2^{-} \circ (1 - F_1)(X_1)$$

almost surely. First,

$$\mathbb{P}(X_2 \le x) = \mathbb{P}(1 - F_1(X_1) \le F_2(x)) = F_2(x).$$

Writing $U = F_1(X_1) \sim \mathrm{Unif}[0,1]$, we obtain

$$\begin{aligned} C(u_1, u_2) &= \mathbb{P}\big(F_1(X_1) \le u_1, F_2(X_2) \le u_2\big) \\ &= \mathbb{P}(U \le u_1, 1 - U \le u_2) \\ &= \max\{u_1 + u_2 - 1, 0\} \\ &= W(u_1, u_2). \end{aligned}$$

Lecture 7 — Wednesday, May 25, 2016

Reverse implication. Suppose that $C = W$, and set $U_j = F_j(X_j)$ for $j \in \{1, 2\}$. By Lemma 6.2, $(U_1, U_2) \sim W$. If $u_1 + u_2 < 1$, then

$$\mathbb{P}(U_1 \le u_1, U_2 \le u_2) = W(u_1, u_2) = 0.$$

If $u_1 + u_2 > 1$, inclusion–exclusion gives

$$\mathbb{P}(U_1 > u_1, U_2 > u_2) = 1 - u_1 - u_2 + W(u_1, u_2) = 0.$$

The two open half-squares on either side of the antidiagonal are countable unions of rectangles with rational endpoints. Consequently,

$$\mathbb{P}(U_1 + U_2 = 1) = 1.$$

Lemma 1.4(3) now yields

$$X_2 = F_2^-(U_2) = F_2^- \circ (1 - F_1)(X_1)$$

almost surely.

It remains to prove (4).

Forward implication. Suppose that

$$X_j = F_j^- \circ F_1(X_1)$$

almost surely for every $j \in \{2, \ldots, d\}$. Since the margins are continuous, $U = F_1(X_1)$ is uniformly distributed on $[0, 1]$, and $F_j(X_j) = U$ almost surely for every j. Hence

$$C(\boldsymbol{u}) = \mathbb{P}(U \le u_1, \ldots, U \le u_d) = \min\{u_1, \ldots, u_d\} = M(\boldsymbol{u}).$$

Reverse implication. Suppose that $C = M$, and again set $U_j = F_j(X_j)$. For each $j \in \{2, \ldots, d\}$ and each $t \in [0, 1]$,

$$\mathbb{P}(U_1 \le t < U_j) = \mathbb{P}(U_1 \le t) - \mathbb{P}(U_1 \le t, U_j \le t) = t - M(t, t) = 0.$$

Interchanging U_1 and U_j gives $\mathbb{P}(U_j \le t < U_1) = 0$. Taking the union over rational $t \in [0, 1]$ shows that $U_j = U_1$ almost surely. Therefore,

$$X_j = F_j^-(U_j) = F_j^- \circ F_1(X_1)$$

almost surely for every $j \in \{2, \ldots, d\}$. □

Remark 7.1 The connection between copulas and dependence was developed systematically by Schweizer and Wolff. Theorem 6.3(1) formalizes the principle that a copula describes dependence independently of the marginal scales. Parts (3) and (4) identify the lower and upper Fréchet–Hoeffding bounds with perfect negative and perfect positive dependence, respectively.

Theorem 7.2 Let $\boldsymbol{X} \sim H$ have margins $F_1, \dots, F_d$ and copula C, which need not be unique. If T_j is measurable, continuous, and strictly increasing on $\operatorname{ran}(X_j)$ for each $j \in \{1, \dots, d\}$, then $(T_1(X_1), \dots, T_d(X_d))$ also has copula C.

Proof. Let $S_j = \operatorname{ran}(X_j)$, which is closed and has full marginal probability, and define

$$q_j(y) = \sup\{x \in S_j : T_j(x) \le y\},$$

with the usual conventions when the set is empty or unbounded. Strict increase makes the sublevel set an initial segment of S_j, while continuity on S_j ensures that a finite boundary point is included when it belongs to S_j. Consequently,

$$\{T_j(X_j) \le y\} = \{X_j \le q_j(y)\}$$

up to a null event. If G_j denotes the distribution function of $T_j(X_j)$, then

$$G_j(y) = F_j(q_j(y)). \tag{7.1}$$

The same event identity applied simultaneously to all coordinates, followed by Theorem 5.2 and (7.1), gives

$$\begin{aligned}
&\mathbb{P}\big(T_1(X_1) \le y_1, \dots, T_d(X_d) \le y_d\big) \\
&\quad = H(q_1(y_1), \dots, q_d(y_d)) \\
&\quad = C\big(F_1(q_1(y_1)), \dots, F_d(q_d(y_d))\big) \\
&\quad = C(G_1(y_1), \dots, G_d(y_d)).
\end{aligned}$$

Therefore C is also a copula of the transformed vector. □

Theorem 7.3 (Survival version of Sklar's theorem) The following statements hold.

1. Let $\overline{H}$ be a joint survival function with marginal survival functions $\overline{F}_1, \dots, \overline{F}_d$. There is a copula $\widehat{C}$ such that

$$\overline{H}(\boldsymbol{x}) = \widehat{C}\big(\overline{F}_1(x_1), \dots, \overline{F}_d(x_d)\big) \tag{7.2}$$

for every $\boldsymbol{x} \in \mathbb{R}^d$. This copula is unique on $\prod_{j=1}^d \operatorname{ran}(\overline{F}_j)$. On that set,

$$\widehat{C}(\boldsymbol{u}) = \overline{H}\big(\overline{F}_1^-(u_1), \dots, \overline{F}_d^-(u_d)\big),$$

where

$$\overline{F}_j^-(u) = \inf\{x \in \mathbb{R} : \overline{F}_j(x) \le u\} = F_j^-(1-u).$$

2. Conversely, let $\widehat{C}$ be a copula and let $\overline{F}_1, \dots, \overline{F}_d$ be univariate survival functions. The function $\overline{H}$ defined by (7.2) is a joint survival function with these marginal survival functions.

Proof. For (1), let C be a copula of the distribution corresponding to $\overline{H}$, whose existence follows from Theorem 5.2. Choose $\boldsymbol{U} \sim C$, and let $\widehat{C}$ be the distribution function of $\mathbf{1} - \boldsymbol{U}$. This is a copula because every coordinate of $\mathbf{1} - \boldsymbol{U}$ is uniformly distributed on $[0,1]$. Inclusion–exclusion and the factorization in Theorem 5.2 then give

$$\begin{aligned}
\overline{H}(\boldsymbol{x}) &= \mathbb{P}(X_1 > x_1, \dots, X_d > x_d) \\
&= \widehat{C}\big(1 - F_1(x_1), \dots, 1 - F_d(x_d)\big) \\
&= \widehat{C}\big(\overline{F}_1(x_1), \dots, \overline{F}_d(x_d)\big).
\end{aligned}$$

The uniqueness and inverse formula follow by the same argument that restricts the range in Theorem 5.2.

For (2), set $F_j = 1 - \overline{F}_j$, choose $\boldsymbol{V} \sim \widehat{C}$, and define

$$X_j = F_j^-(1 - V_j), \quad j \in \{1, \dots, d\}.$$

Each X_j has distribution function F_j. Moreover, the quantile properties in Lemma 1.4 give

$$\begin{aligned}
\mathbb{P}(X_1 > x_1, \dots, X_d > x_d) &= \mathbb{P}\big(V_1 < \overline{F}_1(x_1), \dots, V_d < \overline{F}_d(x_d)\big) \\
&= \widehat{C}\big(\overline{F}_1(x_1), \dots, \overline{F}_d(x_d)\big),
\end{aligned}$$

where strict and nonstrict inequalities agree because the margins of $\boldsymbol{V}$ are continuous. This proves the claim. $\square$

Remark 7.4 The copula $\widehat{C}$ in Theorem 7.3 is the **survival copula** of $\boldsymbol{X}$, or equivalently of its distribution function H. Although $\widehat{C}$ is a distribution function, the functions $\overline{H}, \overline{F}_1, \ldots, \overline{F}_d$ are decreasing rather than increasing and therefore are not distribution functions.

Proposition 7.5 Let C be a copula and $\boldsymbol{U} \sim C$. Then $\mathbf{1} - \boldsymbol{U} \sim \widehat{C}$, where $\widehat{C}$ is the survival copula of $\boldsymbol{U}$. For $J \subseteq \{1, \ldots, d\}$, define

$$a_j(J; \boldsymbol{u}) = \begin{cases} 1 - u_j, & j \in J, \\ 1, & j \notin J. \end{cases}$$

Then

$$\widehat{C}(\boldsymbol{u}) = \sum_{J \subseteq \{1,\ldots,d\}} (-1)^{|J|} C\big(a_1(J; \boldsymbol{u}), \ldots, a_d(J; \boldsymbol{u})\big). \tag{7.3}$$

Proof. Since the margins of $\boldsymbol{U}$ are continuous,

$$\begin{aligned} \widehat{C}(\boldsymbol{u}) &= \mathbb{P}(1 - U_1 \le u_1, \ldots, 1 - U_d \le u_d) \\ &= \mathbb{P}(U_1 > 1 - u_1, \ldots, U_d > 1 - u_d). \end{aligned}$$

Thus $\widehat{C}$ is both the distribution function of $\mathbf{1} - \boldsymbol{U}$ and the survival copula of $\boldsymbol{U}$, by Theorem 7.3. Applying inclusion–exclusion to the final probability gives (7.3). □

Remark 7.6 In two dimensions, (7.3) reduces to

$$\widehat{C}(u_1, u_2) = u_1 + u_2 - 1 + C(1 - u_1, 1 - u_2). \tag{7.4}$$

A copula C is **radially symmetric** when $C = \widehat{C}$. In probabilistic terms, $\boldsymbol{U} \sim C$ then has the same distribution as $\mathbf{1} - \boldsymbol{U}$. Thus, in two dimensions, the distribution is invariant under rotation by 180° about $(1/2, 1/2)$. The copulas W, Π, and M are all radially symmetric, as follows directly from (7.4).

Lecture 8 — Monday, May 30, 2016

3. Copula Properties, Simulation, and Association

Symmetry and Exchangeability

Definition 8.1 A random vector $\boldsymbol{X}$ is **radially symmetric** about $\boldsymbol{a} \in \mathbb{R}^d$ if

$$\boldsymbol{X} - \boldsymbol{a} \overset{\mathrm{d}}{=} \boldsymbol{a} - \boldsymbol{X}.$$

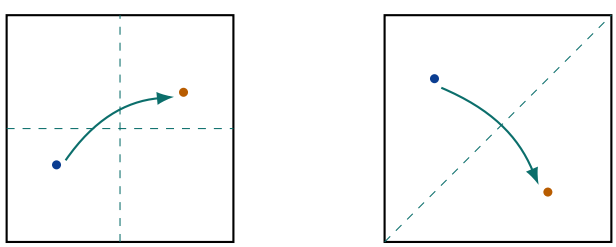
When $d = 1$, this property is simply called symmetry about a.

Definition 8.2 A random vector $\boldsymbol{X}$ is **exchangeable** if

$$(X_{j_1}, \dots, X_{j_d}) \overset{\mathrm{d}}{=} (X_1, \dots, X_d)$$

for every permutation $(j_1, \dots, j_d)$ of $(1, \dots, d)$.

The two symmetries have different geometric meanings. Radial symmetry reflects every coordinate about the centre, whereas exchangeability permutes the coordinate axes. Figure 7 illustrates the distinction in two dimensions.



radial reflection coordinate exchange

Figure 7

Proposition 8.3 Let $\boldsymbol{X} \sim H$ have continuous margins $F_1, \ldots, F_d$ and copula C. The following statements hold.

1. Suppose that X_j is symmetric about a_j for each $j \in \{1, \ldots, d\}$. Then $\boldsymbol{X}$ is radially symmetric about $\boldsymbol{a}$ if and only if $C = \widehat{C}$.
2. The vector $\boldsymbol{X}$ is exchangeable if and only if $F_1 = \cdots = F_d$ and
$$C(u_1, \ldots, u_d) = C(u_{j_1}, \ldots, u_{j_d})$$
for every $\boldsymbol{u} \in [0,1]^d$ and every permutation $(j_1, \ldots, j_d)$ of $(1, \ldots, d)$.

Proof. For (1), set $\boldsymbol{Y} = \boldsymbol{X} - \boldsymbol{a}$, and let G_j be the distribution function of Y_j. By Theorem 7.2, $\boldsymbol{Y}$ has copula C. Marginal symmetry and continuity imply that

$$G_j(-y) = 1 - G_j(y), \quad y \in \mathbb{R}.$$

If $U_j = G_j(Y_j)$, then $\boldsymbol{U} \sim C$ by Lemma 6.2, and the probability transforms of $-\boldsymbol{Y}$ are

$$G_j(-Y_j) = 1 - U_j, \qquad j \in \{1, \ldots, d\}.$$

It follows from Proposition 7.5 that $-\boldsymbol{Y}$ has copula $\widehat{C}$. The vectors $\boldsymbol{Y}$ and $-\boldsymbol{Y}$ have the same continuous margins.

Forward implication. If $\boldsymbol{Y}$ and $-\boldsymbol{Y}$ have the same distribution, the uniqueness assertion in Theorem 5.2 implies that their copulas are equal. Hence $C = \widehat{C}$.

Reverse implication. If $C = \widehat{C}$, then $\boldsymbol{Y}$ and $-\boldsymbol{Y}$ have the same margins and the same copula. Their Sklar factorizations are therefore identical, so $\boldsymbol{Y} \stackrel{\mathrm{d}}{=} -\boldsymbol{Y}$. This proves the first statement.

For (2), fix a permutation $(j_1, \ldots, j_d)$.

Forward implication. Suppose that $\boldsymbol{X}$ is exchangeable. Permutations that interchange two coordinates show that all margins are equal; write $F_1 = \cdots = F_d = F$. Moreover,

$$H(x_{j_1}, \ldots, x_{j_d}) = H(x_1, \ldots, x_d)$$

for every $\boldsymbol{x} \in \mathbb{R}^d$. By the inverse formula in Theorem 5.2,

$$C(u_{j_1}, \ldots, u_{j_d}) = H(F^-(u_{j_1}), \ldots, F^-(u_{j_d}))$$

$$
\begin{aligned}
&= H(F^{-}(u_1), \dots, F^{-}(u_d)) \\
&= C(u_1, \dots, u_d).
\end{aligned}
$$

Reverse implication. Suppose that the margins are all equal to F and that C is invariant under every coordinate permutation. Then

$$H(x_{j_1}, \dots, x_{j_d}) = C(F(x_{j_1}), \dots, F(x_{j_d})) = C(F(x_1), \dots, F(x_d)) = H(x_1, \dots, x_d).$$

Thus $\boldsymbol{X}$ is exchangeable. □

Remark 8.4 The implications among independence, identical marginal distributions, and exchangeability must be distinguished carefully.

1. Identically distributed independent random variables are exchangeable, by Proposition 8.3(2), because their copula is Π. The converse fails: a vector can have equal margins and a permutation-invariant copula different from Π.
2. Exchangeability implies identical marginal distributions, again by Part (2). The converse also fails: equal margins may be combined with a copula that is not invariant under coordinate permutations. A Marshall–Olkin copula with unequal parameters provides such an example.
3. Exchangeability and radial symmetry do not imply one another.

Example 8.5 (Exchangeability without radial symmetry) For $\theta > 0$, the bivariate Clayton copula is

$$C_\theta(u_1, u_2) = \psi_\theta\big(\psi_\theta^{-1}(u_1) + \psi_\theta^{-1}(u_2)\big) \qquad \text{and} \qquad \psi_\theta(t) = (1+t)^{-1/\theta}.$$

This copula is invariant under exchange of u_1 and u_2, so a random vector with uniform margins and copula C_θ is exchangeable. It need not be radially symmetric. For example, when $\theta = 1$,

$$C_1(1/4, 1/4) = \frac{1}{7} \qquad \text{and} \qquad \widehat{C}_1(1/4, 1/4) = -\frac{1}{2} + C_1(3/4, 3/4) = \frac{1}{10}.$$

Thus $C_1 \neq \widehat{C}_1$, so Proposition 8.3(1) shows that the vector is not radially symmetric.

Example 8.6 (Radial symmetry without exchangeability) Let $f_1, f_2 \in C^1[0,1]$ satisfy

$$f_1(0) = f_1(1) = f_2(0) = f_2(1) = 0$$

and

$$1 + f_1'(u_1)f_2'(u_2) \geq 0 \tag{8.1}$$

for every $(u_1, u_2) \in [0,1]^2$. Then

$$C(u_1, u_2) = u_1u_2 + f_1(u_1)f_2(u_2) \tag{8.2}$$

has density

$$c(u_1, u_2) = 1 + f_1'(u_1)f_2'(u_2).$$

Condition (8.1) makes this density nonnegative, and the endpoint conditions give uniform margins. Hence (8.2) defines a copula.

If each f_j is symmetric about $1/2$, then

$$\widehat{C}(u_1, u_2) = u_1 + u_2 - 1 + C(1-u_1, 1-u_2) = u_1u_2 + f_1(1-u_1)f_2(1-u_2) = C(u_1, u_2),$$

where the first equality is (7.4). Thus the copula is radially symmetric. On the other hand,

$$C(u_1, u_2) - C(u_2, u_1) = f_1(u_1)f_2(u_2) - f_1(u_2)f_2(u_1).$$

If f_1 and f_2 are not proportional, this difference is not identically zero, so the copula is not exchangeable.

For a concrete choice, take

$$f_1(u) = \frac{1}{2}u(1-u) \qquad \text{and} \qquad f_2(u) = \frac{1}{2}u(1-u)\big(1 + (u-1/2)^2\big).$$

These functions are symmetric about $1/2$, are not proportional, and satisfy (8.1) because $|f_1'| \leq 1/2$ and $|f_2'| \leq 3/4$.

Sampling Copulas

One motivation from quantitative risk management is the estimation of the value at risk of an aggregate loss. If $S = L_1 + \cdots + L_d$, then

$$\mathrm{VaR}_\alpha(S) := F_S^-(\alpha),$$

often at a high confidence level such as $\alpha = 0.999$. When the marginal loss distributions and their copula are specified separately, the following result provides a direct simulation procedure for their joint distribution.

Proposition 8.7 Let H be a distribution function with margins $F_1, \dots, F_d$ and copula C. If $\boldsymbol{U} \sim C$, then

$$\boldsymbol{X} = \big(F_1^{-}(U_1), \dots, F_d^{-}(U_d)\big) \sim H.$$

Proof. For every $\boldsymbol{x} \in \mathbb{R}^d$, Lemma 1.4(4) and Theorem 5.2 give

$$\begin{aligned}\mathbb{P}(\boldsymbol{X} \le \boldsymbol{x}) &= \mathbb{P}\big(U_1 \le F_1(x_1), \dots, U_d \le F_d(x_d)\big) \\ &= C(F_1(x_1), \dots, F_d(x_d)) \\ &= H(\boldsymbol{x}).\end{aligned}$$

Thus $\boldsymbol{X} \sim H$. □

Proposition 8.8 (Recovery of a copula from continuous margins) Let H be a distribution function with continuous margins $F_1, \dots, F_d$ and copula C. If $\boldsymbol{X} \sim H$, then

$$\big(F_1(X_1), \dots, F_d(X_d)\big) \sim C.$$

Proof. This is the forward implication of Lemma 6.2. Indeed, the marginal probability transforms are uniform by Proposition 2.1, and their joint distribution function is C by the calculation in that lemma. □

Lecture 9 — Wednesday, June 01, 2016

The two transformations in Propositions 8.7 and 8.8 move between a copula and a joint distribution with prescribed margins. A basic example starts with

$$\boldsymbol{X} \sim \mathcal{N}_d(\mathbf{0}, \mathbf{P}),$$

where $\mathbf{P}$ is a correlation matrix. If Φ denotes the standard normal distribution function, then

$$\boldsymbol{U} = \big(\Phi(X_1), \dots, \Phi(X_d)\big)$$

has the **Gaussian copula** with correlation matrix $\mathbf{P}$. Starting from a nondegenerate normal vector $\mathcal{N}_d(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, where $\boldsymbol{\Sigma}$ is positive definite, gives no additional copulas. Indeed, with $\sigma_j = \sqrt{\sigma_{jj}} > 0$, standardizing each coordinate by

$$Y_j = \frac{X_j - \mu_j}{\sigma_j}$$

is a strictly increasing transformation, so Theorem 6.3 shows that $\boldsymbol{X}$ and $\boldsymbol{Y}$ have the same copula.

To simulate an arbitrary copula recursively, conditional distributions replace the explicit normal construction. Let C be a d-dimensional copula and let $\boldsymbol{U} \sim C$. A version of the regular conditional distribution function of U_j given $(U_1, \dots, U_{j-1})$ is denoted by

$$C(u_j \mid u_1, \dots, u_{j-1}) \coloneqq \mathbb{P}\big(U_j \le u_j \mid U_1 = u_1, \dots, U_{j-1} = u_{j-1}\big). \tag{9.1}$$

Because the event on the right usually has probability zero, this expression is understood as a regular conditional distribution and is determined only almost everywhere with respect to the conditioning distribution.

For indices $j_1, \dots, j_k$, write

$$C^{(j_1,\dots,j_k)}(u_{j_1}, \dots, u_{j_k}) = C(v_1, \dots, v_d),$$

where $v_j = u_j$ when $j \in \{j_1, \dots, j_k\}$ and $v_j = 1$ otherwise. Thus $C^{(j_1,\dots,j_k)}$ is the corresponding marginal copula.

Theorem 9.1 Fix $j \in \{2, \dots, d\}$. Suppose that the marginal copulas $C^{(1,\dots,j)}$ and $C^{(1,\dots,j-1)}$ are absolutely continuous in their first $j-1$ arguments and admit continuous mixed partial derivatives with respect to those arguments. At every point where the denominator below is positive,

$$C(u_j \mid u_1, \dots, u_{j-1}) = \frac{D_{1,\dots,j-1} C^{(1,\dots,j)}(u_1, \dots, u_j)}{D_{1,\dots,j-1} C^{(1,\dots,j-1)}(u_1, \dots, u_{j-1})} \tag{9.2}$$

for almost every $(u_1, \dots, u_{j-1}) \in [0,1]^{j-1}$.

Proof. Let $\boldsymbol{U}_{1:j-1} = (U_1, \dots, U_{j-1})$. The joint density of $\boldsymbol{U}_{1:j-1}$ is

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

whereas the density of the subprobability measure

$$A \longmapsto \mathbb{P}(\boldsymbol{U}_{1:j-1} \in A, U_j \le u_j)$$

is

$$D_{1,\dots,j-1} C^{(1,\dots,j)}(u_1,\dots,u_j).$$

The conditional distribution in (9.1) is the Radon–Nikodym derivative of the second measure with respect to the first. Taking the ratio of these densities gives (9.2). This formula also appears in Schmitz (2003, p. 20). □

The identity for conditional distributions

$$F_{X_1,X_2}(x_1,x_2) = \int_{-\infty}^{x_1} F_{X_2|X_1}(x_2 \mid \widetilde{x}_1)\, \mathrm{d}F_{X_1}(\widetilde{x}_1) \tag{9.3}$$

is the bivariate instance of disintegration. It reconstructs a joint distribution by averaging the conditional distribution over the first coordinate.

Theorem 9.2 (Conditional distribution method) Let C be a d-dimensional copula, and let $U_1',\dots,U_d'$ be independent random variables uniformly distributed on $[0,1]$. Choose measurable versions of the regular conditional distribution functions and their generalized inverses, and define recursively

$$\begin{aligned} U_1 &= U_1', \\ U_j &= C^{-}(U_j' \mid U_1,\dots,U_{j-1}), \quad j \in \{2,\dots,d\}, \end{aligned}$$

where

$$C^{-}(v \mid u_1,\dots,u_{j-1}) = \inf\{t \in [0,1] : C(t \mid u_1,\dots,u_{j-1}) \ge v\}.$$

Then $\boldsymbol{U} = (U_1,\dots,U_d) \sim C$.

Proof. We prove by induction on j that $(U_1,\dots,U_j)$ has marginal copula $C^{(1,\dots,j)}$. The claim is immediate for $j=1$.

For the bivariate step, the quantile transformation in Proposition 2.1 and the independence of U_1' and U_2' give

$$\mathbb{P}(U_1 \le u_1, U_2 \le u_2) = \int_0^{u_1} \mathbb{P}\big(U_2' \le C(u_2 \mid \widetilde{u}_1)\big)\, \mathrm{d}\widetilde{u}_1$$

$$
\begin{aligned}
&= \int_0^{u_1} C(u_2 \mid \widetilde{u}_1)\,\mathrm{d}\widetilde{u}_1 \\
&= C(u_1, u_2),
\end{aligned}
$$

where the final equality is the identity for conditional distributions in (9.3).

Write

$$\boldsymbol{U}_{1:j-1} = (U_1, \dots, U_{j-1}) \qquad \text{and} \qquad \widetilde{\boldsymbol{u}}_{1:j-1} = (\widetilde{u}_1, \dots, \widetilde{u}_{j-1}).$$

Suppose that $\boldsymbol{U}_{1:j-1}$ has distribution function $C^{(1,\dots,j-1)}$. Conditional on $\boldsymbol{U}_{1:j-1} = \widetilde{\boldsymbol{u}}_{1:j-1}$, the same quantile argument gives

$$\mathbb{P}(U_j \le u_j \mid U_1 = \widetilde{u}_1, \dots, U_{j-1} = \widetilde{u}_{j-1}) = C(u_j \mid \widetilde{u}_1, \dots, \widetilde{u}_{j-1}).$$

Consequently,

$$
\begin{aligned}
&\mathbb{P}(U_1 \le u_1, \dots, U_j \le u_j) \\
&\quad = \int_{[0,u_1]\times\cdots\times[0,u_{j-1}]} C(u_j \mid \widetilde{u}_1, \dots, \widetilde{u}_{j-1})\,\mathrm{d}C^{(1,\dots,j-1)}(\widetilde{u}_1, \dots, \widetilde{u}_{j-1}) \\
&\quad = C^{(1,\dots,j)}(u_1, \dots, u_j),
\end{aligned}
$$

again by disintegration. This completes the induction. □

Remark 9.3 The conditional distribution method and its inverse have the following roles.

1. The construction in Theorem 9.2 transforms independent uniform random variables into a vector with copula C. It is the multivariate analogue of the inverse transform construction in Proposition 2.1.
2. The method applies broadly, but it may require repeated numerical evaluation and inversion of conditional distribution functions. For a particular copula family, a specialized stochastic representation may be faster and more numerically stable.
3. When the relevant conditional distribution functions are continuous and strictly increasing, the inverse map is the **Rosenblatt transformation**

   $$R(\boldsymbol{u}) = \big(u_1, C(u_2 \mid u_1), \dots, C(u_d \mid u_1, \dots, u_{d-1})\big). \tag{9.4}$$

   If $\boldsymbol{U} \sim C$, then $R(\boldsymbol{U})$ has independent uniform coordinates. This fact supports goodness-of-fit testing: after fitting a copula, we can test the transformed observations for uniformity and independence.

Example 9.4 Consider the bivariate upper Fréchet–Hoeffding bound $M(u_1, u_2) = \min\{u_1, u_2\}$.

Solution. Fix $u_1 \in (0,1)$. By Theorem 9.1, a version of the conditional distribution function is

$$M(u_2 \mid u_1) = D_1 M(u_1, u_2) = \begin{cases} 0, & u_2 < u_1, \\ 1, & u_2 \geq u_1. \end{cases} \tag{9.5}$$

The derivative fails to exist only at $u_2 = u_1$, where the right-continuous version in (9.5) assigns the value 1. The conditional distribution is therefore concentrated at u_1, as shown in Figure 8.

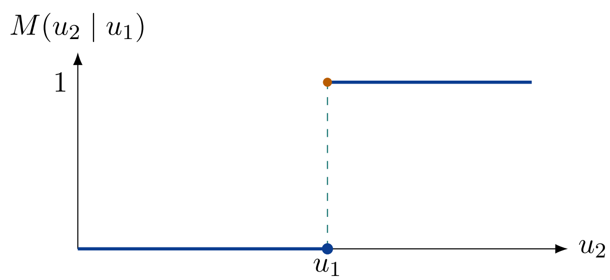


Figure 8

Let U_1' and U_2' be independent and uniformly distributed on $[0,1]$. The generalized inverse of (9.5) is

$$M^{-}(v \mid u_1) = \inf\{u_2 \in [0,1] : M(u_2 \mid u_1) \geq v\} = u_1$$

for every $v \in (0,1]$. Consequently, the construction in Theorem 9.2 gives

$$(U_1, U_2) = \big(U_1', M^{-}(U_2' \mid U_1')\big) = \big(U_1', U_1'\big).$$

This is precisely the comonotone representation of M from Theorem 6.3(4). $\square$

Lecture 10 — Monday, June 06, 2016

Measures of Association

Example 10.1 Consider the d-dimensional Clayton copula

$$C(\boldsymbol{u}) = \psi_\theta\big(\psi_\theta^{-1}(u_1) + \cdots + \psi_\theta^{-1}(u_d)\big) \qquad \text{and} \qquad \psi_\theta(t) = (1+t)^{-1/\theta},$$

where $\theta > 0$. Determine the conditional distribution functions required by Theorem 9.2 and their generalized inverses.

Solution. The generator and its derivatives satisfy

$$\psi_\theta^{-1}(u) = u^{-\theta} - 1$$

and

$$\psi_\theta^{(k)}(t) = (-1)^k \prod_{\ell=0}^{k-1} \left(\ell + \frac{1}{\theta}\right) (1+t)^{-(k+1/\theta)} \tag{10.1}$$

for every positive integer k. Its typical shape is shown in Figure 9.

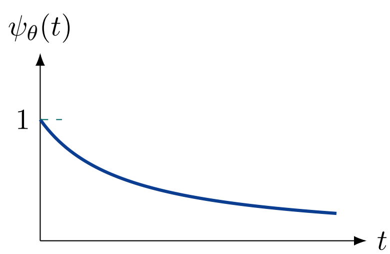


Figure 9

Applying Theorem 9.1 and then (10.1) gives, for $j \in \{2, \ldots, d\}$,

$$C(u_j \mid u_1, \ldots, u_{j-1}) = \frac{\psi_\theta^{(j-1)}\left(\sum_{k=1}^{j} \psi_\theta^{-1}(u_k)\right)}{\psi_\theta^{(j-1)}\left(\sum_{k=1}^{j-1} \psi_\theta^{-1}(u_k)\right)}$$

$$= \left(\frac{1 - j + \sum_{k=1}^{j} u_k^{-\theta}}{2 - j + \sum_{k=1}^{j-1} u_k^{-\theta}} \right)^{-(j-1+1/\theta)}.$$

Set

$$A_{j-1} = 2 - j + \sum_{k=1}^{j-1} u_k^{-\theta}.$$

Solving $v = C(u_j \mid u_1, \dots, u_{j-1})$ for u_j yields

$$C^{-}(v \mid u_1, \dots, u_{j-1}) = \left[1 + A_{j-1} \left(v^{-1/(j-1+1/\theta)} - 1\right)\right]^{-1/\theta}. \tag{10.2}$$

Equation (10.2), used recursively, gives a direct implementation of the conditional distribution method. □

Example 10.2 Let V be positive with Laplace transform

$$\psi(t) = \mathbb{E}[e^{-tV}] = \int_0^\infty e^{-tv} f_V(v)\,\mathrm{d}v. \tag{10.3}$$

Let $E_1, \dots, E_d$ be independent standard exponential random variables, independent of V, and define

$$U_j = \psi(E_j/V), \quad j \in \{1, \dots, d\}.$$

Show that $\boldsymbol{U}$ has Archimedean copula

$$C(\boldsymbol{u}) = \psi\big(\psi^{-1}(u_1) + \cdots + \psi^{-1}(u_d)\big).$$

Solution. Since ψ is decreasing, conditioning on V gives

$$\begin{aligned} \mathbb{P}(U_1 \le u_1, \dots, U_d \le u_d \mid V) &= \mathbb{P}\big(E_j \ge V\psi^{-1}(u_j) \text{ for every } j \mid V\big) \\ &= \exp\left(-V \sum_{j=1}^{d} \psi^{-1}(u_j)\right). \end{aligned}$$

Taking expectations and using (10.3) yields

$$\mathbb{P}(\boldsymbol{U} \le \boldsymbol{u}) = \psi\left(\sum_{j=1}^{d} \psi^{-1}(u_j)\right) = C(\boldsymbol{u}).$$

For the Clayton generator $\psi_\theta(t) = (1+t)^{-1/\theta}$, the mixing variable has the gamma distribution with shape $1/\theta$ and rate 1. $\square$

Definition 10.3 The **Fréchet class** $\mathscr{F}(F_1, \ldots, F_d)$ is the set of all distribution functions with margins $F_1, \ldots, F_d$. By Theorem 5.2,

$$\mathscr{F}(F_1, \ldots, F_d) = \{H : H(\boldsymbol{x}) = C(F_1(x_1), \ldots, F_d(x_d)) \text{ for some copula } C\}.$$

Thus a Fréchet class fixes the marginal distributions while allowing the dependence structure to vary.

In applications, it is often useful to summarize the degree of dependence between two random variables by a real number. Before introducing scalar measures of association, it is helpful to compare copulas through a partial order.

Definition 10.4 For two d-dimensional copulas C_1 and C_2, write

$$C_1 \preceq C_2$$

if $C_1(\boldsymbol{u}) \le C_2(\boldsymbol{u})$ for every $\boldsymbol{u} \in [0,1]^d$. This relation is the **lower-orthant order**. When $d = 2$, it is also the concordance order: the bivariate identity

$$\widehat{C}(v_1, v_2) = v_1 + v_2 - 1 + C(1 - v_1, 1 - v_2)$$

shows that the same inequality automatically holds for the survival copulas. Accordingly, in the bivariate setting used below, $C_1 \preceq C_2$ means that C_1 is less concordant than C_2.

Definition 10.5 Let X_1 and X_2 have finite, positive variances. Their **Pearson correlation coefficient** is

$$\begin{aligned}\rho(X_1, X_2) &= \frac{\operatorname{Cov}(X_1, X_2)}{\sqrt{\operatorname{Var}(X_1)\operatorname{Var}(X_2)}} \\ &= \frac{\mathbb{E}\big[(X_1 - \mathbb{E}[X_1])(X_2 - \mathbb{E}[X_2])\big]}{\sqrt{\mathbb{E}[(X_1 - \mathbb{E}[X_1])^2]\mathbb{E}[(X_2 - \mathbb{E}[X_2])^2]}}.\end{aligned}$$

For observations (x_{i1}, x_{i2}), $i \in \{1, \ldots, n\}$, from a random sample, the **sample Pearson correlation coefficient** is

$$\widehat{\rho}_n = \frac{\sum_{i=1}^n (x_{i1} - \overline{x}_1)(x_{i2} - \overline{x}_2)}{\sqrt{\sum_{i=1}^n (x_{i1} - \overline{x}_1)^2 \sum_{i=1}^n (x_{i2} - \overline{x}_2)^2}}, \tag{10.4}$$

where

$$\overline{x}_j = \frac{1}{n}\sum_{i=1}^{n} x_{ij}, \quad j \in \{1,2\}.$$

The expression in (10.4) is defined whenever neither sample variance is zero.

Proposition 10.6 (Hoeffding's covariance identity) Let $(X_1, X_2) \sim H$ have margins F_1, F_2 and finite second moments. Then

$$\operatorname{Cov}(X_1, X_2) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \big(H(x_1,x_2) - F_1(x_1)F_2(x_2)\big)\,\mathrm{d}x_1\,\mathrm{d}x_2. \tag{10.5}$$

Proof. Let (X_1', X_2') be an independent copy of (X_1, X_2). Independence between the primed and unprimed vectors gives

$$\mathbb{E}\big[(X_1 - X_1')(X_2 - X_2')\big] = 2\operatorname{Cov}(X_1, X_2). \tag{10.6}$$

For real numbers a and b,

$$a - b = \int_{-\infty}^{\infty} \big(\mathbb{1}_{\{b \le x\}} - \mathbb{1}_{\{a \le x\}}\big)\,\mathrm{d}x. \tag{10.7}$$

Applying (10.7) to both factors in (10.6), and then applying Fubini's theorem, gives

$$\begin{aligned}
&2\operatorname{Cov}(X_1, X_2)\\
&= \int_{\mathbb{R}^2} \mathbb{E}\Big[\big(\mathbb{1}_{\{X_1' \le x_1\}} - \mathbb{1}_{\{X_1 \le x_1\}}\big)\big(\mathbb{1}_{\{X_2' \le x_2\}} - \mathbb{1}_{\{X_2 \le x_2\}}\big)\Big]\,\mathrm{d}\boldsymbol{x}.
\end{aligned}$$

The expectation in the integrand equals

$$\begin{aligned}
&\mathbb{P}(X_1' \le x_1, X_2' \le x_2) - \mathbb{P}(X_1' \le x_1)\mathbb{P}(X_2 \le x_2)\\
&\quad - \mathbb{P}(X_1 \le x_1)\mathbb{P}(X_2' \le x_2) + \mathbb{P}(X_1 \le x_1, X_2 \le x_2)\\
&= 2\big(H(x_1,x_2) - F_1(x_1)F_2(x_2)\big).
\end{aligned}$$

Substitution and cancellation of the factor 2 prove (10.5). The use of Fubini's theorem is justified because the finite second moments imply

$$\mathbb{E}\big[|X_1 - X_1'||X_2 - X_2'|\big] < \infty$$

by the Cauchy–Schwarz inequality. □

Lecture 11 — Wednesday, June 08, 2016

Proposition 11.1 Let X_1 and X_2 have finite, positive variances. The following statements hold.

1. We have $|\rho(X_1, X_2)| \leq 1$. Equality holds if and only if
$$X_2 = a + bX_1$$
almost surely for some $a, b \in \mathbb{R}$ with $b \neq 0$.
2. If X_1 and X_2 are independent, then $\rho(X_1, X_2) = 0$.
3. Pearson correlation is invariant under strictly increasing affine transformations of both variables.
4. For $i, j \in \{1, 2\}$, let X_{ij} have distribution function F_j, and let (X_{i1}, X_{i2}) have copula C_i. If $C_1 \preceq C_2$, then
$$\rho(X_{11}, X_{12}) \leq \rho(X_{21}, X_{22}).$$

Proof. For (1), set $Y_j = X_j - \mathbb{E}[X_j]$. The space L^2 is a Hilbert space under the inner product $\langle Y, Z\rangle = \mathbb{E}[YZ]$. The Cauchy–Schwarz inequality gives

$$\begin{aligned} |\operatorname{Cov}(X_1, X_2)|^2 &= |\langle Y_1, Y_2\rangle|^2 \\ &\leq \langle Y_1, Y_1\rangle\langle Y_2, Y_2\rangle \\ &= \operatorname{Var}(X_1)\operatorname{Var}(X_2). \end{aligned}$$

Dividing by the positive product of the variances proves $|\rho(X_1, X_2)| \leq 1$.

Forward implication. Suppose that $|\rho(X_1, X_2)| = 1$. Equality holds in the Cauchy–Schwarz inequality, so $Y_2 = bY_1$ almost surely for some nonzero b. Therefore,

$$X_2 = \mathbb{E}[X_2] - b\mathbb{E}[X_1] + bX_1,$$

which has the required affine form.

Reverse implication. Suppose that $X_2 = a + bX_1$ almost surely with $b \neq 0$. Then

$$\rho(X_1, X_2) = \frac{b}{|b|},$$

so $|\rho(X_1, X_2)| = 1$. In particular, positive and negative slopes correspond to correlations 1 and -1, respectively.

For (2), independence gives

$$\begin{aligned}
\operatorname{Cov}(X_1, X_2) &= \mathbb{E}\big[(X_1 - \mathbb{E}[X_1])(X_2 - \mathbb{E}[X_2])\big] \\
&= \mathbb{E}[X_1 - \mathbb{E}[X_1]]\mathbb{E}[X_2 - \mathbb{E}[X_2]] \\
&= 0.
\end{aligned}$$

Hence $\rho(X_1, X_2) = 0$.

For (3), let $a_1, a_2 > 0$. Then

$$\begin{aligned}
&\operatorname{Cov}(a_1X_1 + b_1, a_2X_2 + b_2) = a_1a_2 \operatorname{Cov}(X_1, X_2), \\
&\operatorname{Var}(a_jX_j + b_j) = a_j^2 \operatorname{Var}(X_j), \quad j \in \{1, 2\}.
\end{aligned}$$

Substitution into Definition 10.5 proves the invariance.

Finally, consider (4). By Proposition 10.6 and Theorem 5.2,

$$\operatorname{Cov}(X_{i1}, X_{i2}) = \int_{\mathbb{R}^2} \big(C_i(F_1(x_1), F_2(x_2)) - F_1(x_1)F_2(x_2)\big)\,\mathrm{d}\boldsymbol{x}.$$

The assumption $C_1 \preceq C_2$ makes the first integrand no larger than the second. Thus

$$\operatorname{Cov}(X_{11}, X_{12}) \leq \operatorname{Cov}(X_{21}, X_{22}).$$

The two pairs have the same marginal variances, so division by their common product of positive standard deviations proves the desired correlation ordering. □

Remark 11.2 The following common claims are false.

1. Pearson correlation does not exist for every pair of random variables. For example, a Pareto random variable with tail index 3/2 has an infinite second moment, so it cannot be assigned a Pearson correlation with another variable through Definition 10.5.

2. Pearson correlation is not invariant under arbitrary strictly increasing transformations. Let X_1 and X_2 be independent and identically distributed Pareto random variables with tail index 3, so that
$$F(x) = 1 - x^{-3}, \quad x \geq 1.$$
Proposition 11.1(2) gives $\rho(X_1, X_2) = 0$. The transformation $x \mapsto x^2$ is strictly increasing on their range, but X_1^2 has a Pareto distribution with tail index $3/2$. Indeed,
$$\mathbb{P}(X_1^2 > x) = \mathbb{P}(X_1 > \sqrt{x}) = x^{-3/2}, \quad x \geq 1,$$
and therefore $\mathrm{Var}(X_1^2) = \infty$. Consequently, $\rho(X_1^2, X_2)$ is not defined.

3. Zero Pearson correlation does not imply independence.

4. Marginal distribution functions together with Pearson correlation do not determine the joint distribution function.

The last two claims fail simultaneously within the copula family from Example 8.6. Let
$$\begin{aligned} f_1(u) &= \frac{1}{2}u(1-u)(1-2u), \\ f_2(u) &= \frac{1}{2}u(1-u), \end{aligned}$$
and define
$$C(u_1, u_2) = u_1 u_2 + f_1(u_1) f_2(u_2). \tag{11.1}$$
The derivative bounds $|f_1'| \leq 1/2$ and $|f_2'| \leq 1/2$ ensure that $1 + f_1'(u_1) f_2'(u_2) \geq 3/4$, so (11.1) is a copula. If $(U_1, U_2) \sim C$, then both margins are uniform and hence have variance $1/12$. By Proposition 10.6,
$$\begin{aligned} \rho(U_1, U_2) &= 12 \,\mathrm{Cov}(U_1, U_2) \\ &= 12 \int_0^1 \int_0^1 \big(C(u_1, u_2) - u_1 u_2\big) \,\mathrm{d}u_1 \,\mathrm{d}u_2 \\ &= 12 \left(\int_0^1 f_1(u_1) \,\mathrm{d}u_1 \right) \left(\int_0^1 f_2(u_2) \,\mathrm{d}u_2 \right) \\ &= 0, \end{aligned}$$
because $f_1(1-u) = -f_1(u)$. Nevertheless, $C \neq \Pi$, so the variables are dependent. Moreover, varying the nonzero multiplier in front of $f_1 f_2$, while retaining a nonnegative density, produces distinct joint distributions with the same uniform margins and the same zero correlation.

Lecture 12 — Monday, June 13, 2016

The counterexample in (11.1) belongs to a full one-parameter family. Set

$$f_1(u) = 2u(u - 1/2)(u - 1) \qquad \text{and} \qquad f_{2,\theta}(u) = \theta u(1 - u), \quad \theta \in [-1, 1].$$

The function f_1 is antisymmetric about 1/2, so its integral over $[0, 1]$ is zero. Moreover, $|f_1'| \leq 1$ and $|f_{2,\theta}'| \leq 1$, which makes

$$C_\theta(u_1, u_2) = u_1 u_2 + f_1(u_1) f_{2,\theta}(u_2)$$

a copula. The calculation in Remark 11.2 shows that every C_θ has zero Pearson correlation. For nonzero distinct values of θ, these are distinct dependent joint models with the same uniform margins.

The same construction extends to nonuniform margins with finite second moments. Indeed, Proposition 10.6 and Theorem 5.2 make the covariance contribution factor as

$$\left(\int_{\mathbb{R}} f_1(F_1(x))\, \mathrm{d}x \right) \left(\int_{\mathbb{R}} f_{2,\theta}(F_2(y))\, \mathrm{d}y \right).$$

If F_1 is symmetric about zero, then $F_1(-x) = 1 - F_1(x)$, and the first integrand is odd. Its integral vanishes, so the resulting variables remain uncorrelated without being independent.

There is a further limitation: for fixed margins, not every value in $[-1, 1]$ need be attainable as a Pearson correlation. To see this, suppose that

$$X_j = e^{\sigma_j Z_j} \qquad \text{and} \qquad Z_j \sim \mathcal{N}(0, 1), \quad j \in \{1, 2\}.$$

Thus X_j has a lognormal distribution with parameters $(0, \sigma_j^2)$. Parts (4) and (3)– (4) imply that the smallest and largest correlations occur at W and M, respectively.

For the countermonotone coupling, take

$$(X_1, X_2) = (e^{\sigma_1 Z}, e^{-\sigma_2 Z}) \qquad \text{and} \qquad Z \sim \mathcal{N}(0, 1).$$

Since $\mathbb{E}[e^{tZ}] = e^{t^2/2}$,

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

Using

$$\mathrm{Var}(e^{\sigma Z}) = (e^{\sigma^2} - 1)e^{\sigma^2},$$

we obtain

$$\widetilde{\rho}(s,t) = \frac{e^{st} - 1}{\sqrt{(e^{s^2} - 1)(e^{t^2} - 1)}}. \tag{12.1}$$

Therefore,

$$\rho_{\min} = \widetilde{\rho}(\sigma_1, -\sigma_2) \qquad \text{and} \qquad \rho_{\max} = \widetilde{\rho}(\sigma_1, \sigma_2). \tag{12.2}$$

The upper endpoint follows from the same calculation with the comonotone coupling $(e^{\sigma_1 Z}, e^{\sigma_2 Z})$. For example, if $\sigma_1^2 = 1$ and $\sigma_2^2 = 16$, then

$$-0.0003 \lesssim \rho \lesssim 0.0137.$$

Thus even the entire Fréchet class of these margins permits only a very narrow range of Pearson correlations.

The preceding examples expose three principal drawbacks of Pearson correlation.

1. It is not defined for every pair of random variables.
2. Its attainable values and numerical magnitude depend on the marginal distributions.
3. It is invariant under increasing affine transformations, but not under all increasing transformations.

To address these drawbacks, Scarsini (1984) proposed an axiomatic definition.

Definition 12.1 A **measure of concordance** is a functional $\kappa(X_1, X_2)$ on pairs of continuously distributed random variables that satisfies the following axioms.

1. The value $\kappa(X_1, X_2)$ is defined for every such pair.
2. The range of κ is $[-1, 1]$.
3. We have $\kappa(X_1, X_2) = \kappa(X_2, X_1)$.
4. We have $\kappa(-X_1, X_2) = -\kappa(X_1, X_2)$.
5. Independence of X_1 and X_2 implies $\kappa(X_1, X_2) = 0$.

6. If (X_{i1}, X_{i2}) has copula C_i for $i \in \{1, 2\}$ and $C_1 \preceq C_2$, then
$$\kappa(X_{11}, X_{12}) \leq \kappa(X_{21}, X_{22}).$$
7. Suppose that $(X_{n1}, X_{n2}) \sim H_n$ for every positive integer n, that $(X_1, X_2) \sim H$, and that the distribution functions are continuous. If H_n converges pointwise to H, then
$$\kappa(X_{n1}, X_{n2}) \longrightarrow \kappa(X_1, X_2).$$

Proposition 12.2 Let κ be a measure of concordance, and let (X_1, X_2) be a pair of continuously distributed random variables with copula C. The following statements hold.

1. The value of $\kappa(X_1, X_2)$ depends only on C, and may therefore be written as $\kappa(C)$.
2. If T_j is strictly increasing on $\operatorname{ran}(X_j)$ for $j \in \{1, 2\}$, then
$$\kappa(T_1(X_1), T_2(X_2)) = \kappa(X_1, X_2).$$
3. If T_1 is strictly increasing on $\operatorname{ran}(X_1)$ and T_2 is strictly decreasing on $\operatorname{ran}(X_2)$, then
$$\kappa(T_1(X_1), T_2(X_2)) = -\kappa(X_1, X_2).$$
4. The Fréchet–Hoeffding bounds satisfy $\kappa(W) = -1$ and $\kappa(M) = 1$.

Proof. For (1), let $\boldsymbol{U} \sim C$. The copulas of both $\boldsymbol{X}$ and $\boldsymbol{U}$ are C. Applying Definition 12.1(6) in both directions gives
$$\kappa(X_1, X_2) \leq \kappa(U_1, U_2) \qquad \text{and} \qquad \kappa(U_1, U_2) \leq \kappa(X_1, X_2).$$
Hence the two values are equal and depend only on C.

For (2), Theorem 6.3(1) shows that the transformed pair has the same copula C. The conclusion follows from (1).

For (3), Definition 12.1(3) and (4) imply that
$$\kappa(Y_1, -Y_2) = -\kappa(Y_1, Y_2)$$

for every continuously distributed pair (Y_1, Y_2). Since $-T_2$ is strictly increasing, Part (2) gives

$$\kappa(T_1(X_1), T_2(X_2)) = -\kappa(T_1(X_1), -T_2(X_2)) = -\kappa(X_1, X_2).$$

Finally, Theorem 4.2 and the ordering axiom give

$$\kappa(W) \leq \kappa(C) \leq \kappa(M)$$

for every bivariate copula C. The range axiom supplies copulas at which the values -1 and 1 are attained. The displayed inequalities and the range restriction then force $\kappa(W) = -1$ and $\kappa(M) = 1$. $\square$

Remark 12.3 The axioms address several, but not all, limitations of Pearson correlation. They ensure existence for every continuously distributed pair and invariance under arbitrary strictly increasing marginal transformations. They do not make zero concordance equivalent to independence. For example, let (X_1, X_2) be uniformly distributed on the unit circle, as illustrated in Figure 10.

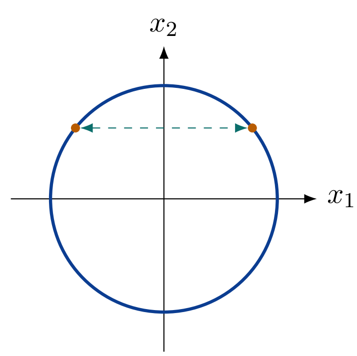


Figure 10

The coordinates are dependent because $X_1^2 + X_2^2 = 1$ almost surely, but

$$(X_1, X_2) \stackrel{\mathrm{d}}{=} (-X_1, X_2).$$

The reflection axiom therefore gives

$$\kappa(X_1, X_2) = \kappa(-X_1, X_2) = -\kappa(X_1, X_2),$$

and hence $\kappa(X_1, X_2) = 0$.

Nor do the margins and a single concordance value determine the joint law. For instance, distinct Archimedean families such as the Clayton and Gumbel families can be calibrated to have the same Kendall's tau.

The full range $[-1, 1]$, however, is attainable for fixed continuous margins. For $t \in [0, 1]$, define

$$C_t = (1 - t)W + tM.$$

This is a copula by Remark 4.1. The map $t \mapsto \kappa(C_t)$ is continuous by Definition 12.1(7), and its endpoint values are -1 and 1 by Proposition 12.2(4). The intermediate value theorem therefore gives every value in $[-1, 1]$.

Two particularly important measures of concordance are Spearman's rho and Kendall's tau.

Definition 12.4 Let (X_1, X_2) have continuous margins F_1, F_2, and let (X_1', X_2') be an independent copy. **Spearman's rho** is

$$\rho_S(X_1, X_2) = \rho(F_1(X_1), F_2(X_2)),$$

and **Kendall's tau** is

$$\tau(X_1, X_2) = \mathbb{E}\left[\operatorname{sign}\big((X_1 - X_1')(X_2 - X_2')\big)\right],$$

where

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

Thus Spearman's rho is the Pearson correlation of the probability transforms, whereas Kendall's tau is the probability of concordance minus the probability of discordance between two independent observations.

Proposition 12.5 Let (X_1, X_2) be continuously distributed with copula C. Then the following formulas hold.

1. We have

$$\rho_S(X_1, X_2) = 12 \int_0^1 \int_0^1 \big(C(u_1, u_2) - u_1 u_2\big)\, \mathrm{d}u_1\, \mathrm{d}u_2$$

$$= 12 \int_0^1 \int_0^1 C(u_1, u_2)\,\mathrm{d}u_1\,\mathrm{d}u_2 - 3.$$

2. If $(U_1, U_2) \sim C$, then

$$\tau(X_1, X_2) = 4 \int_{[0,1]^2} C(u_1, u_2)\,\mathrm{d}C(u_1, u_2) - 1 = 4\mathbb{E}[C(U_1, U_2)] - 1.$$

Proof. For (1), let $U_j = F_j(X_j)$. Each U_j is uniformly distributed on $[0, 1]$, so $\mathbb{E}[U_j] = 1/2$ and $\mathrm{Var}(U_j) = 1/12$. By Proposition 10.6,

$$\begin{aligned} \rho_S(X_1, X_2) &= 12\,\mathrm{Cov}(U_1, U_2) \\ &= 12 \int_0^1 \int_0^1 \big(C(u_1, u_2) - u_1 u_2\big)\,\mathrm{d}u_1\,\mathrm{d}u_2. \end{aligned}$$

Since the integral of $u_1 u_2$ over the unit square is $1/4$, the second formula follows.

For (2), continuity of the margins makes ties have probability zero. Hence

$$\tau(X_1, X_2) = 2\mathbb{P}\big((X_1 - X_1')(X_2 - X_2') > 0\big) - 1 = 4\mathbb{P}(X_1 > X_1', X_2 > X_2') - 1,$$

where the second equality uses the symmetry between the two independent observations. Conditioning on (X_1, X_2) gives

$$\begin{aligned} \tau(X_1, X_2) &= 4\mathbb{E}\big[\mathbb{P}(X_1' < X_1, X_2' < X_2 \mid X_1, X_2)\big] - 1 \\ &= 4\mathbb{E}[H(X_1, X_2)] - 1 \\ &= 4\mathbb{E}[C(F_1(X_1), F_2(X_2))] - 1 \\ &= 4 \int_{[0,1]^2} C(u_1, u_2)\,\mathrm{d}C(u_1, u_2) - 1. \end{aligned}$$

□

Example 12.6 Let (X_1, X_2) have a Gaussian copula with correlation parameter $r \in [-1, 1]$. Then

$$\rho_S(X_1, X_2) = \frac{6}{\pi} \arcsin\left(\frac{r}{2}\right) \qquad \text{and} \qquad \tau(X_1, X_2) = \frac{2}{\pi} \arcsin(r).$$

Solution. We use the bivariate normal quadrant identity

$$\mathbb{P}(Z_1 > 0, Z_2 > 0) = \frac{1}{4} + \frac{1}{2\pi}\arcsin(q), \tag{12.3}$$

where (Z_1, Z_2) is standard bivariate normal with correlation q.

Let (Z_1, Z_2) be standard bivariate normal with correlation r, and let Y_1, Y_2 be independent standard normal variables, independent of (Z_1, Z_2). Then

$$\begin{aligned}\mathbb{E}[\Phi(Z_1)\Phi(Z_2)] &= \mathbb{P}(Y_1 \le Z_1, Y_2 \le Z_2)\\ &= \mathbb{P}(Z_1 - Y_1 \ge 0, Z_2 - Y_2 \ge 0).\end{aligned}$$

The two differences have correlation $r/2$. Applying (12.3) and then Proposition 12.5(1) yields

$$\rho_S = 12\left(\frac{1}{4} + \frac{1}{2\pi}\arcsin(r/2)\right) - 3 = \frac{6}{\pi}\arcsin(r/2).$$

For Kendall's tau, let (Z_1', Z_2') be an independent copy. The vector $(Z_1 - Z_1', Z_2 - Z_2')$ is bivariate normal with correlation r. Therefore,

$$\begin{aligned}\tau &= 4\mathbb{P}(Z_1 - Z_1' > 0, Z_2 - Z_2' > 0) - 1\\ &= 4\left(\frac{1}{4} + \frac{1}{2\pi}\arcsin(r)\right) - 1\\ &= \frac{2}{\pi}\arcsin(r),\end{aligned}$$

again by (12.3). $\square$

Lecture 13 — Wednesday, June 15, 2016

Suppose that (x_{i1}, x_{i2}), $i \in \{1, \ldots, n\}$, are observations from a continuously distributed random vector. Let R_{ij} be the rank of x_{ij} among $x_{1j}, \ldots, x_{nj}$. With probability one there are no ties, and the sample version of Spearman's rho is the sample Pearson correlation of the two rank vectors:

$$\widehat{\rho}_{S,n} = \frac{\sum_{i=1}^n (R_{i1} - \overline{R}_{\cdot 1})(R_{i2} - \overline{R}_{\cdot 2})}{\sqrt{\sum_{i=1}^n (R_{i1} - \overline{R}_{\cdot 1})^2 \sum_{i=1}^n (R_{i2} - \overline{R}_{\cdot 2})^2}}$$

$$= \frac{12}{n(n^2-1)} \left[\sum_{i=1}^{n} R_{i1} R_{i2} - n \left(\frac{n+1}{2} \right)^2 \right]. \tag{13.1}$$

Indeed,

$$\overline{R}_{\cdot j} = \frac{n+1}{2} \qquad \text{and} \qquad \sum_{i=1}^{n} (R_{ij} - \overline{R}_{\cdot j})^2 = \frac{n(n^2-1)}{12},$$

because each rank vector is a permutation of $(1, \dots, n)$.

The sample version of Kendall's tau is

$$\widehat{\tau}_n = \frac{1}{\binom{n}{2}} \sum_{1 \le i < j \le n} \operatorname{sign}\big((x_{i1} - x_{j1})(x_{i2} - x_{j2})\big). \tag{13.2}$$

It is the proportion of concordant sample pairs minus the proportion of discordant sample pairs.

Lemma 13.1 For any bivariate copulas C and C',

$$\int_{[0,1]^2} C(\boldsymbol{u}) \, \mathrm{d}C'(\boldsymbol{u}) = \int_{[0,1]^2} C'(\boldsymbol{u}) \, \mathrm{d}C(\boldsymbol{u}). \tag{13.3}$$

Proof. Let $(U_1, U_2) \sim C$ and $(U_1', U_2') \sim C'$ be independent. By conditioning on the primed vector,

$$\mathbb{P}(U_1 \le U_1', U_2 \le U_2') = \int_{[0,1]^2} C(u_1, u_2) \, \mathrm{d}C'(u_1, u_2).$$

The continuous margins make coordinate ties have probability zero. Hence the event on the left agrees almost surely with $\{U_1' > U_1, U_2' > U_2\}$. Conditioning instead on (U_1, U_2) and using inclusion–exclusion gives

$$\begin{aligned}
&\mathbb{P}(U_1' > U_1, U_2' > U_2) \\
&\quad = \int_{[0,1]^2} \big(1 - u_1 - u_2 + C'(u_1, u_2)\big) \, \mathrm{d}C(u_1, u_2) \\
&\quad = \int_{[0,1]^2} C'(u_1, u_2) \, \mathrm{d}C(u_1, u_2),
\end{aligned}$$

because $\mathbb{E}[U_1] = \mathbb{E}[U_2] = 1/2$. This proves (13.3). $\square$

Taking $C' = \Pi$ in Lemma 13.1 yields

$$\int_{[0,1]^2} u_1 u_2 \, \mathrm{d}C(u_1, u_2) = \int_0^1 \int_0^1 C(u_1, u_2) \, \mathrm{d}u_1 \, \mathrm{d}u_2.$$

Consequently, the lemma gives an alternative derivation of the Spearman formula in (1); the Kendall formula in (2) is the other part of the same proposition.

Proposition 13.2 Let (X_1, X_2) have continuous marginal distribution functions and copula C. If κ denotes either Spearman's rho or Kendall's tau, then the following statements hold.

1. If $\kappa = 1$, then $C = M$.
2. If $\kappa = -1$, then $C = W$.

Proof. First suppose that $\kappa = \rho_S$. By Proposition 12.5(1),

$$\rho_S = 12 \int_{[0,1]^2} C(\boldsymbol{u}) \, \mathrm{d}\boldsymbol{u} - 3.$$

If $\rho_S = 1$, then $\int C \, \mathrm{d}\boldsymbol{u} = 1/3$, which is also the integral of M. Since $C \leq M$ and both functions are continuous, equality of their integrals implies $C = M$. Similarly, if $\rho_S = -1$, then $\int C \, \mathrm{d}\boldsymbol{u} = 1/6 = \int W \, \mathrm{d}\boldsymbol{u}$. The inequality $C \geq W$ and continuity therefore imply $C = W$.

Now suppose that $\kappa = \tau$, and let $(U_1, U_2) \sim C$. Proposition 12.5(2) gives

$$\tau = 4\mathbb{E}[C(U_1, U_2)] - 1.$$

If $\tau = 1$, then $\mathbb{E}[C(U_1, U_2)] = 1/2$. The inequalities

$$C(U_1, U_2) \leq \min\{U_1, U_2\} \leq U_j, \quad j \in \{1, 2\},$$

and $\mathbb{E}[U_j] = 1/2$ show that $U_1 = \min\{U_1, U_2\} = U_2$ almost surely. Thus the vector is comonotone, and its copula is M. If $\tau = -1$, then $\mathbb{E}[C(U_1, U_2)] = 0$. Since

$$C(U_1, U_2) \geq W(U_1, U_2) \geq 0,$$

we have $W(U_1, U_2) = 0$ almost surely, so $U_1 + U_2 \leq 1$ almost surely. But $\mathbb{E}[U_1 + U_2] = 1$, and consequently $U_1 + U_2 = 1$ almost surely. The vector is countermonotone, and its copula is W. $\square$

Remark 13.3 If a copula C is absolutely continuous and has sufficiently regular first partial derivatives, then integration by parts gives

$$\int_{[0,1]^2} C(\boldsymbol{u})\,\mathrm{d}C(\boldsymbol{u}) = \frac{1}{2} - \int_{[0,1]^2} D_1C(\boldsymbol{u})D_2C(\boldsymbol{u})\,\mathrm{d}\boldsymbol{u}. \tag{13.4}$$

Indeed, writing the copula density as $D_{12}C$, integration by parts in the first coordinate produces the boundary term $\int_0^1 u_2\,\mathrm{d}u_2 = 1/2$. Formula (13.4) is often useful for computing Kendall's tau; see Nelsen (2006) for further details.

Tail Dependence

Tail dependence measures the strength of association when both components of a bivariate random vector are simultaneously in the lower tail or simultaneously in the upper tail. The two regions are illustrated in Figure 11.

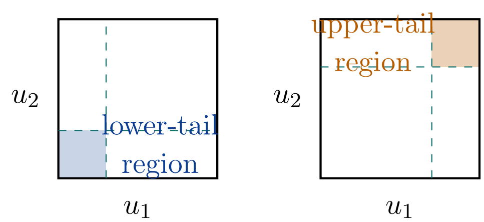


Figure 11

Definition 13.4 Let (X_1, X_2) have continuous marginal distribution functions F_1 and F_2. Provided that the limits exist, the **lower coefficient of tail dependence** and the **upper coefficient of tail dependence** are, respectively,

$$\lambda_L = \lim_{u\downarrow 0} \mathbb{P}\big(X_2 \le F_2^-(u) \mid X_1 \le F_1^-(u)\big), \tag{13.5}$$

$$\lambda_U = \lim_{u\uparrow 1} \mathbb{P}\big(X_2 > F_2^-(u) \mid X_1 > F_1^-(u)\big). \tag{13.6}$$

Proposition 13.5 Let (X_1, X_2) have continuous marginal distribution functions and copula C, and write $\delta_C(u) = C(u,u)$. Then the following statements hold.

1. For every $u \in (0,1]$,

$$\mathbb{P}\big(X_2 \leq F_2^-(u) \mid X_1 \leq F_1^-(u)\big) = \frac{C(u,u)}{u}. \tag{13.7}$$

   Thus λ_L exists if and only if the following limit exists, in which case

$$\lambda_L = \lim_{u \downarrow 0} \frac{\delta_C(u)}{u}. \tag{13.8}$$

2. If δ_C is differentiable near zero and the indicated limit exists, then

$$\lambda_L = \lim_{u \downarrow 0} \delta_C'(u). \tag{13.9}$$

   If, in addition, C is totally differentiable near $(0,0)$, then

$$\lambda_L = \lim_{u \downarrow 0} \big(D_1 C(u,u) + D_2 C(u,u)\big), \tag{13.10}$$

   whenever this limit exists.

3. For every $u \in [0,1)$,

$$\begin{aligned}\mathbb{P}\big(X_2 > F_2^-(u) \mid X_1 > F_1^-(u)\big) \\ = \frac{1-2u+C(u,u)}{1-u} = \frac{\widehat{C}(1-u,1-u)}{1-u}.\end{aligned} \tag{13.11}$$

   Thus λ_U exists if and only if either of the following equivalent limits exists, in which case

$$\lambda_U = \lim_{u \uparrow 1} \frac{1-2u+\delta_C(u)}{1-u} = \lim_{v \downarrow 0} \frac{\widehat{C}(v,v)}{v}. \tag{13.12}$$

4. If δ_C is differentiable near one and the indicated limit exists, then

$$\lambda_U = 2 - \lim_{u \uparrow 1} \delta_C'(u). \tag{13.13}$$

   If, in addition, C is totally differentiable near $(1,1)$, then

$$\lambda_U = 2 - \lim_{u \uparrow 1} \big(D_1 C(u,u) + D_2 C(u,u)\big), \tag{13.14}$$

   whenever this limit exists.

Proof. Set $U_j = F_j(X_j)$ for $j \in \{1,2\}$. Continuity of the marginal distribution functions implies that $(U_1, U_2) \sim C$ and that both components are uniformly distributed. Therefore, for $u \in (0,1]$,

$$\begin{aligned}&\mathbb{P}\big(X_2 \le F_2^-(u) \mid X_1 \le F_1^-(u)\big)\\ &\qquad = \mathbb{P}(U_2 \le u \mid U_1 \le u) = \frac{\mathbb{P}(U_1 \le u, U_2 \le u)}{\mathbb{P}(U_1 \le u)} = \frac{C(u,u)}{u}.\end{aligned}$$

This proves Part (1). Since $\delta_C(0) = 0$, l'Hôpital's rule gives (13.9) under the stated assumptions. The chain rule yields

$$\delta_C'(u) = D_1 C(u,u) + D_2 C(u,u), \tag{13.15}$$

which completes Part (2).

For the upper tail, inclusion–exclusion gives

$$\begin{aligned}\mathbb{P}(U_1 > u, U_2 > u) &= 1 - \mathbb{P}(U_1 \le u) - \mathbb{P}(U_2 \le u) + \mathbb{P}(U_1 \le u, U_2 \le u)\\ &= 1 - 2u + C(u,u).\end{aligned}$$

Dividing by $\mathbb{P}(U_1 > u) = 1 - u$ proves the first equality in (13.11). The second follows from the bivariate survival-copula identity $\widehat{C}(v,v) = 2v - 1 + C(1-v, 1-v)$, established in Theorem 7.3. This proves Part (3). Finally, l'Hôpital's rule gives

$$\lim_{u \uparrow 1} \frac{1 - 2u + \delta_C(u)}{1-u} = 2 - \lim_{u \uparrow 1} \delta_C'(u),$$

and combining this identity with (13.15) proves Part (4). □

The limits in Definition 13.4 need not exist. The following notion isolates the function of one variable that controls them.

Definition 13.6 A function $\delta\colon [0,1] \to [0,1]$ is a **bivariate copula diagonal** if it satisfies the following conditions.

1. We have $\delta(1) = 1$.
2. We have $\delta(u) \le u$ for every $u \in [0,1]$.
3. For all $0 \le u \le v \le 1$,
$$0 \le \delta(v) - \delta(u) \le 2(v-u). \tag{13.16}$$

If C is a bivariate copula, then the function $\delta_C(u) = C(u, u)$ is called the **diagonal section** of C. Parts (2) and (3) together imply $\delta(0) = 0$. The endpoint ratios in (13.8) and (13.12) can oscillate, so the corresponding tail dependence coefficient can fail to exist.

Lecture 14 — Monday, June 20, 2016

Proposition 14.1 The diagonal section of every bivariate copula is a bivariate copula diagonal in the sense of Definition 13.6.

Proof. Let $\delta_C(u) = C(u, u)$. The copula margins give $\delta_C(1) = C(1, 1) = 1$, and monotonicity gives

$$\delta_C(u) = C(u, u) \le C(u, 1) = u$$

for every $u \in [0, 1]$. If $0 \le u \le v \le 1$, monotonicity and the Lipschitz bound in Remark 4.1(1) yield

$$0 \le C(v, v) - C(u, u) \le 2(v - u).$$

Thus all three conditions in Definition 13.6 are satisfied. □

The converse also holds: every function satisfying those three elementary conditions occurs as the diagonal section of a copula.

Theorem 14.2 Let δ be a bivariate copula diagonal. Then

$$C_\delta(u_1, u_2) = \min\left\{u_1, u_2, \frac{\delta(u_1) + \delta(u_2)}{2}\right\} \tag{14.1}$$

is a copula whose diagonal section is δ.

Proof. Define

$$F_1(t) = \delta(t) \qquad \text{and} \qquad F_2(t) = 2t - \delta(t), \quad t \in [0, 1].$$

The observations following Definition 13.6 give $\delta(0) = 0$. Hence $F_1(0) = F_2(0) = 0$, while $F_1(1) = F_2(1) = 1$. Part (3) of that definition shows that, whenever $0 \le s \le t \le 1$,

$$F_1(t) - F_1(s) = \delta(t) - \delta(s) \ge 0,$$

$$F_2(t) - F_2(s) = 2(t-s) - \big(\delta(t) - \delta(s)\big) \geq 0.$$

Thus F_1 and F_2 are distribution functions supported on $[0,1]$.

Now define

$$C'(u_1,u_2) = \frac{1}{2}\left[M\big(F_1(u_1), F_2(u_2)\big) + M\big(F_2(u_1), F_1(u_2)\big)\right]. \tag{14.2}$$

Each summand is a bivariate distribution function, so their convex mixture is also a bivariate distribution function. Moreover,

$$\begin{aligned} C'(u,1) &= \frac{F_1(u) + F_2(u)}{2} = u, \\ C'(1,u) &= \frac{F_2(u) + F_1(u)}{2} = u. \end{aligned}$$

Therefore, C' is a copula. It remains to verify that $C' = C_\delta$.

Fix $(u_1,u_2) \in [0,1]^2$, and set

$$m = \frac{\delta(u_1) + \delta(u_2)}{2}.$$

The inequalities

$$\begin{aligned} 2u_1 - \delta(u_1) &\geq \delta(u_2), \\ 2u_2 - \delta(u_2) &\geq \delta(u_1) \end{aligned}$$

are equivalent to $u_1 \geq m$ and $u_2 \geq m$, respectively. We consider the four possible combinations.

If $u_1 \geq m$ and $u_2 \geq m$, then

$$C'(u_1,u_2) = \frac{\delta(u_1) + \delta(u_2)}{2} = m = C_\delta(u_1,u_2).$$

If $u_1 \geq m$ and $u_2 \leq m$, then

$$C'(u_1,u_2) = \frac{2u_2 - \delta(u_2) + \delta(u_2)}{2} = u_2 = C_\delta(u_1,u_2).$$

The case $u_1 \leq m$ and $u_2 \geq m$ is symmetric and gives $C'(u_1,u_2) = u_1 = C_\delta(u_1,u_2)$. Finally, suppose that both $u_1 \leq m$ and $u_2 \leq m$. Since $\delta(u_j) \leq u_j$,

$$u_1 + u_2 \leq 2m = \delta(u_1) + \delta(u_2) \leq u_1 + u_2.$$

Equality holds throughout, so $u_1 = u_2 = m$, and again $C'(u_1,u_2) = m = C_\delta(u_1,u_2)$. This proves $C' = C_\delta$.

Finally, Definition 13.6(2) gives

$$C_\delta(u,u) = \min\{u, \delta(u)\} = \delta(u),$$

so C_δ has the required diagonal section. □

The construction in Theorem 14.2 allows a diagonal to oscillate between the curves u^2 and u increasingly rapidly near one. This produces a copula without an upper coefficient of tail dependence.

Corollary 14.3 There exists a bivariate copula for which λ_U does not exist.

Proof. Let $a_1 = 1/2$, and recursively define

$$b_n = 2a_n - a_n^2 \qquad \text{and} \qquad a_{n+1} = \sqrt{b_n}, \quad n \in \mathbb{N}.$$

Both sequences increase to one. Define $\delta(u) = u^2$ on $[0, a_1]$, and, for every $n \in \mathbb{N}$, extend it linearly by

$$\delta(u) = \begin{cases} a_n^2 + 2(u - a_n), & a_n \le u \le b_n, \\ b_n, & b_n \le u \le a_{n+1}. \end{cases} \tag{14.3}$$

The pieces meet because $a_n^2 + 2(b_n - a_n) = b_n$ and $a_{n+1}^2 = b_n$. The function is nondecreasing, has slopes between zero and two, lies below the identity, and satisfies $\delta(1) = 1$. Hence it is a bivariate copula diagonal. Its first few pieces are shown in Figure 12.

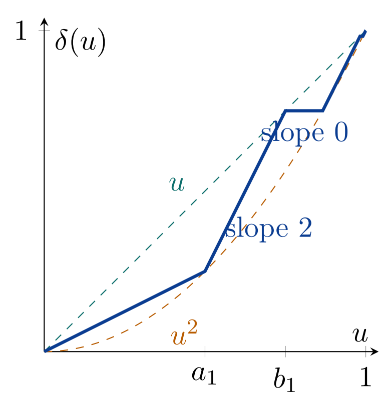


Figure 12

Let C_δ be the copula supplied by Theorem 14.2. At the alternating endpoints, $\delta(a_n) = a_n^2$ and

$\delta(b_n) = b_n$. Consequently,

$$\lim_{n\to\infty} \frac{1-2a_n+C_\delta(a_n,a_n)}{1-a_n} = \lim_{n\to\infty}(1-a_n) = 0,$$
$$\lim_{n\to\infty} \frac{1-2b_n+C_\delta(b_n,b_n)}{1-b_n} = 1.$$

Thus the limit in (13.12) does not exist, and neither does λ_U. □

Example 14.4 For $\theta > 0$, compute λ_L and λ_U for the Clayton copula C_θ.

Solution. On the diagonal,

$$C_\theta(u,u) = \big(2u^{-\theta}-1\big)^{-1/\theta} = u\big(2-u^{\theta}\big)^{-1/\theta}.$$

Proposition 13.5(1) therefore gives

$$\lambda_L = \lim_{u\downarrow 0} \frac{C_\theta(u,u)}{u} = 2^{-1/\theta}.$$

Moreover, differentiating the diagonal gives

$$\frac{\mathrm{d}}{\mathrm{d}u}C_\theta(u,u) = 2u^{-\theta-1}\big(2u^{-\theta}-1\big)^{-1/\theta-1},$$

whose limit as $u \uparrow 1$ is 2. Proposition 13.5(4) yields $\lambda_U = 2-2 = 0$. □

Extensions to Higher Dimensions

Let $(X_1,\dots,X_d)$ have continuous marginal distribution functions and copula C, and write $U_j = F_j(X_j)$. Several extensions of bivariate tail dependence have been proposed. In this subsection, $d \geq 2$.

Definition 14.5 Provided that the indicated limits exist, the following quantities measure different aspects of joint tail dependence.

1. Let J be a nonempty proper subset of $\{1,\dots,d\}$, let C^J be the marginal copula of $(U_j)_{j\in J}$, and let $\widehat{C}^J$ be its survival copula. The lower and upper coefficients relative to

the partition (J, J^{c}) are

$$\begin{aligned}
\lambda_{L,J} &= \lim_{u\downarrow 0} \mathbb{P}\left(X_k \le F_k^-(u) \text{ for all } k \notin J \;\middle|\; X_j \le F_j^-(u) \text{ for all } j \in J\right) \\
&= \lim_{u\downarrow 0} \frac{C(u,\dots,u)}{C^J(u,\dots,u)}, \qquad (14.4) \\
\lambda_{U,J} &= \lim_{u\uparrow 1} \mathbb{P}\left(X_k > F_k^-(u) \text{ for all } k \notin J \;\middle|\; X_j > F_j^-(u) \text{ for all } j \in J\right) \\
&= \lim_{u\uparrow 1} \frac{\widehat{C}(1-u,\dots,1-u)}{\widehat{C}^J(1-u,\dots,1-u)}. \qquad (14.5)
\end{aligned}$$

The copula is lower or upper tail dependent relative to this partition if the corresponding coefficient is positive.

2. The lower and upper **extremal dependence coefficients** introduced by Frahm (2006) are

$$\begin{aligned}
\varepsilon_L &= \lim_{u\downarrow 0} \mathbb{P}\left(\max_{1\le j\le d} U_j \le u \;\middle|\; \min_{1\le j\le d} U_j \le u\right) \\
&= \lim_{u\downarrow 0} \frac{C(u,\dots,u)}{1-\widehat{C}(1-u,\dots,1-u)}, \qquad (14.6) \\
\varepsilon_U &= \lim_{u\uparrow 1} \mathbb{P}\left(\min_{1\le j\le d} U_j > u \;\middle|\; \max_{1\le j\le d} U_j > u\right) \\
&= \lim_{u\uparrow 1} \frac{\widehat{C}(1-u,\dots,1-u)}{1-C(u,\dots,u)}. \qquad (14.7)
\end{aligned}$$

The denominators are the probabilities that at least one component enters the relevant tail, whereas the numerators are the probabilities that all components enter that tail.

3. Schmid and Schmidt (2007) introduced the conditional version of Spearman's rho for the lower tail

$$\rho_{S,L}(u) = \frac{\displaystyle\int_{[0,u]^d} C(\boldsymbol{v})\,\mathrm{d}\boldsymbol{v} - \left(\frac{u^2}{2}\right)^d}{\dfrac{u^{d+1}}{d+1} - \left(\dfrac{u^2}{2}\right)^d}, \quad u \in (0,1]. \qquad (14.8)$$

The subtracted term is the integral of the independence copula over $[0,u]^d$, and the first term in the denominator is the corresponding integral of the upper Fréchet–Hoeffding bound. The limit in the lower tail is

$$\rho_{S,L} = \lim_{u \downarrow 0} \rho_{S,L}(u). \tag{14.9}$$

Lecture 15 — Wednesday, June 22, 2016

4. Elliptical Copulas

Elliptical Distributions and Copulas

Elliptical distributions arise by applying an affine transformation to a random point whose direction is uniform on the unit sphere and whose distance from the origin is governed by an independent radial variable.

Definition 15.1 Let

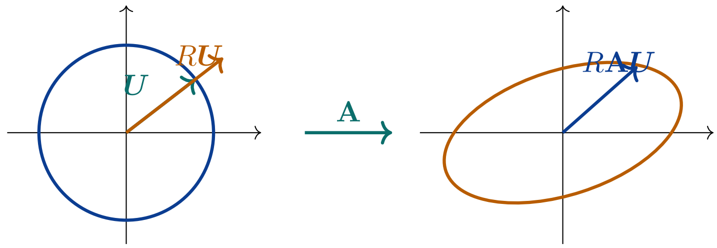
$$S_d = \{\boldsymbol{x} \in \mathbb{R}^d : \|\boldsymbol{x}\| = 1\}.$$

A random vector $\boldsymbol{X} = (X_1, \dots, X_d)$ has an **elliptical distribution** with location vector $\boldsymbol{\mu} \in \mathbb{R}^d$, positive-definite scale matrix $\boldsymbol{\Sigma} \in \mathbb{R}^{d\times d}$, and nonnegative radial variable R if

$$\boldsymbol{X} \overset{\mathrm{d}}{=} \boldsymbol{\mu} + R\mathbf{A}\boldsymbol{U}, \tag{15.1}$$

where $\mathbf{A}\mathbf{A}^\mathsf{T} = \boldsymbol{\Sigma}$, $\boldsymbol{U}$ is uniformly distributed on S_d, and R and $\boldsymbol{U}$ are independent. We can take $\mathbf{A}$ to be the Cholesky factor of $\boldsymbol{\Sigma}$. We write $\boldsymbol{X} \sim E_d(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \phi)$, where ϕ is the characteristic generator introduced below.

The geometry of (15.1) is shown in Figure 13. The radius R selects a sphere, the direction $\boldsymbol{U}$ selects a point on that sphere, and the linear map $\mathbf{A}$ changes spherical level sets into ellipses.



spherical representation elliptical representation

Figure 13

Theorem 15.2 Suppose that the setting of Definition 15.1 holds. Then the following statements are valid.

1. The characteristic function of $\boldsymbol{X}$ is

$$\Phi_{\boldsymbol{X}}(\boldsymbol{t}) = \mathbb{E}\left[e^{i\boldsymbol{t}^\top \boldsymbol{X}}\right] = e^{i\boldsymbol{t}^\top \boldsymbol{\mu}}\phi(\boldsymbol{t}^\top \boldsymbol{\Sigma}\boldsymbol{t}), \tag{15.2}$$

where

$$\phi(s) = \int_0^\infty \Omega_d(r^2 s)\,\mathrm{d}F_R(r). \tag{15.3}$$

Here F_R is the distribution function of R, and Ω_d is determined by

$$\mathbb{E}\left[e^{i\boldsymbol{s}^\top \boldsymbol{U}}\right] = \Omega_d(\boldsymbol{s}^\top \boldsymbol{s}).$$

2. The vector $\boldsymbol{X}$ has a density if and only if R has a density. In that case there is a density generator $g\colon [0,\infty) \to [0,\infty)$ for which

$$f_{\boldsymbol{X}}(\boldsymbol{x}) = \frac{1}{\sqrt{\det(\boldsymbol{\Sigma})}} g\left((\boldsymbol{x}-\boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})\right). \tag{15.4}$$

The radial density and the density generator are related by

$$f_R(r) = \frac{2\pi^{d/2}}{\Gamma(d/2)} r^{d-1} g(r^2), \qquad r > 0, \tag{15.5}$$

$$f_{R^2}(s) = \frac{\pi^{d/2}}{\Gamma(d/2)} s^{d/2-1} g(s), \qquad s > 0. \tag{15.6}$$

3. If $d \geq 2$ and $\boldsymbol{X}$ has a density, then $X_j \sim E_1(\mu_j, \sigma_{jj}, g_1)$, where
$$g_1(t) = \frac{\pi^{(d-1)/2}}{\Gamma((d-1)/2)} \int_t^\infty (y-t)^{(d-1)/2-1} g(y)\,\mathrm{d}y, \quad t \geq 0. \tag{15.7}$$
In particular,
$$F_j(x) = \frac{1}{2} + \frac{\pi^{(d-1)/2}}{\Gamma((d-1)/2)\sqrt{\sigma_{jj}}} \int_{\mu_j}^x \int_{(y-\mu_j)^2/\sigma_{jj}}^\infty \left(z - \frac{(y-\mu_j)^2}{\sigma_{jj}}\right)^{(d-1)/2-1} g(z)\,\mathrm{d}z\,\mathrm{d}y. \tag{15.8}$$
Consequently, the standardized components $(X_j - \mu_j)/\sqrt{\sigma_{jj}}$ are identically distributed.
4. If $\boldsymbol{\mu} = \mathbf{0}$ and $\mathbf{A} = \mathbf{I}_d$, so that the distribution is spherical, then on the event $\{R > 0\}$,
$$\left(\|\boldsymbol{X}\|, \frac{\boldsymbol{X}}{\|\boldsymbol{X}\|}\right) = (R, \boldsymbol{U}). \tag{15.9}$$
5. If $\mathbb{E}[R^2] < \infty$, then
$$\mathrm{Cov}(\boldsymbol{X}) = \frac{\mathbb{E}[R^2]}{d}\boldsymbol{\Sigma}. \tag{15.10}$$
If $\mathbb{E}[R^2] > 0$, the component variances are positive, and the correlation matrix $\mathbf{P} = (\rho_{ij})_{i,j=1}^d$ is determined by
$$\rho_{ij} = \frac{\sigma_{ij}}{\sqrt{\sigma_{ii}\sigma_{jj}}}. \tag{15.11}$$

Proof. For Part (1), condition first on $R = r$. Since $\left\|\mathbf{A}^\mathsf{T}\boldsymbol{t}\right\|^2 = \boldsymbol{t}^\mathsf{T}\boldsymbol{\Sigma}\boldsymbol{t}$, rotational invariance of $\boldsymbol{U}$ gives
$$\begin{aligned}\mathbb{E}\left[e^{i\boldsymbol{t}^\mathsf{T}\boldsymbol{X}} \mid R = r\right] &= e^{i\boldsymbol{t}^\mathsf{T}\boldsymbol{\mu}}\mathbb{E}\left[e^{ir(\mathbf{A}^\mathsf{T}\boldsymbol{t})^\mathsf{T}\boldsymbol{U}}\right] \\ &= e^{i\boldsymbol{t}^\mathsf{T}\boldsymbol{\mu}}\Omega_d\left(r^2\boldsymbol{t}^\mathsf{T}\boldsymbol{\Sigma}\boldsymbol{t}\right).\end{aligned}$$
Integrating with respect to F_R proves (15.2) and (15.3).

To prove Part (2), first consider the spherical vector $R\boldsymbol{U}$. Polar integration shows that a density of the form $g(\|\boldsymbol{x}\|^2)$ induces the radial density
$$f_R(r) = \mathrm{area}(S_d) r^{d-1} g(r^2) = \frac{2\pi^{d/2}}{\Gamma(d/2)} r^{d-1} g(r^2).$$
Conversely, this formula defines g from any radial density. The affine change of variables $\boldsymbol{x} =$

$\boldsymbol{\mu} + \mathbf{A}\boldsymbol{y}$, whose Jacobian has absolute determinant $\sqrt{\det(\boldsymbol{\Sigma})}$, then gives (15.4). The usual change of variables in one dimension from R to R^2 gives (15.6).

For Part (3), standardize one component and integrate the spherical density over the remaining $d-1$ coordinates. Polar coordinates in $\mathbb{R}^{d-1}$ give

$$\int_{\mathbb{R}^{d-1}} g\big(t + \|\boldsymbol{y}\|^2\big)\,\mathrm{d}\boldsymbol{y} \\ = \frac{\pi^{(d-1)/2}}{\Gamma((d-1)/2)} \int_t^\infty (s-t)^{(d-1)/2-1} g(s)\,\mathrm{d}s,$$

which is (15.7). Scaling and integrating this marginal density from its symmetry centre μ_j yields (15.8).

Part (4) follows immediately from $\|R\boldsymbol{U}\| = R$ and $\|\boldsymbol{U}\| = 1$. Finally, rotational invariance gives $\mathbb{E}[\boldsymbol{U}] = \mathbf{0}$ and $\mathbb{E}[\boldsymbol{U}\boldsymbol{U}^\top] = \mathbf{I}_d/d$. Independence of R and $\boldsymbol{U}$ therefore implies

$$\operatorname{Cov}(\boldsymbol{X}) = \mathbf{A}\mathbb{E}\left[R^2\boldsymbol{U}\boldsymbol{U}^\top\right]\mathbf{A}^\top = \frac{\mathbb{E}[R^2]}{d}\mathbf{A}\mathbf{A}^\top,$$

proving Part (5). These properties are also developed in McNeil, Frey, and Embrechts (2005). □

Remark 15.3 Suppose that $\boldsymbol{X} \sim E_d(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \phi)$, and let $\mathbf{D} = \operatorname{diag}(\sqrt{\sigma_{11}}, \ldots, \sqrt{\sigma_{dd}})$. Assume that the margins of $\boldsymbol{X}$ are continuous; when $d \geq 2$, it is enough to assume $\mathbb{P}(R = 0) = 0$.

1. The standardized vector

$$\boldsymbol{Y} = \mathbf{D}^{-1}(\boldsymbol{X} - \boldsymbol{\mu}) \sim E_d(\mathbf{0}, \mathbf{P}, \phi) \qquad \text{and} \qquad \mathbf{P} = \mathbf{D}^{-1}\boldsymbol{\Sigma}\mathbf{D}^{-1},$$

   has the same copula as $\boldsymbol{X}$, by Theorem 6.3(1). Thus an elliptical copula is determined by the correlation matrix $\mathbf{P}$ and the radial law.

2. Since $\boldsymbol{U} \stackrel{\mathrm{d}}{=} -\boldsymbol{U}$,

$$\boldsymbol{X} - \boldsymbol{\mu} \stackrel{\mathrm{d}}{=} R\mathbf{A}\boldsymbol{U} \stackrel{\mathrm{d}}{=} -R\mathbf{A}\boldsymbol{U} \stackrel{\mathrm{d}}{=} \boldsymbol{\mu} - \boldsymbol{X}. \tag{15.12}$$

   Therefore, every elliptical copula is radially symmetric, so $C = \widehat{C}$. In two dimensions, every correlation matrix has the form $\left(\begin{smallmatrix} 1 & \rho \\ \rho & 1 \end{smallmatrix}\right)$, which is invariant under coordinate exchange; hence every bivariate elliptical copula is exchangeable. In higher dimensions, full exchangeability corresponds to the equicorrelation form

$$\mathbf{P} = (1-\rho)\mathbf{I}_d + \rho\mathbf{1}\mathbf{1}^\mathsf{T},$$

subject to the usual positive-definiteness constraint $-1/(d-1) < \rho < 1$.

Proposition 15.4 Let $\mathbf{P}$ be a positive-definite correlation matrix, let $\mathbf{A}$ satisfy $\mathbf{A}\mathbf{A}^\mathsf{T} = \mathbf{P}$, and let F be the common marginal distribution function of $R\mathbf{A}\boldsymbol{U}$. Assume that F is continuous; for $d \geq 2$, it is enough to require $\mathbb{P}(R = 0) = 0$. The following procedure produces one observation from the associated elliptical copula.

1. Sample R from F_R.
2. Independently generate $\boldsymbol{Z} \sim \mathcal{N}_d(\mathbf{0}, \mathbf{I}_d)$, and set $\boldsymbol{U} = \boldsymbol{Z}/\|\boldsymbol{Z}\|$.
3. Set $\boldsymbol{X} = R\mathbf{A}\boldsymbol{U}$.
4. Return $(F(X_1), \dots, F(X_d))$.

Proof. Rotational invariance of the standard multivariate normal distribution implies that $\boldsymbol{Z}/\|\boldsymbol{Z}\|$ is uniformly distributed on S_d and is independent of $\|\boldsymbol{Z}\|$. Thus (3) has the stochastic representation in Definition 15.1. Because $\mathbf{P}$ has unit diagonal, Theorem 15.2(3) shows that all components have the common distribution function F. The probability transform in Part (4) therefore preserves the copula and produces uniform margins. □

Lemma 15.5 Let $\boldsymbol{X} \sim E_2(\mathbf{0}, \boldsymbol{\Sigma}, \phi)$, suppose that $\mathbb{P}(\boldsymbol{X} = \mathbf{0}) = 0$, and let

$$\rho = \frac{\sigma_{12}}{\sqrt{\sigma_{11}\sigma_{22}}}.$$

Then

$$\mathbb{P}(X_1 > 0, X_2 > 0) = \frac{1}{4} + \frac{1}{2\pi}\arcsin(\rho). \tag{15.13}$$

Proof. After positive coordinatewise rescaling, which does not change signs, it is enough to consider $\boldsymbol{Y} \sim E_2(\mathbf{0}, \mathbf{P}, \phi)$, where

$$\mathbf{P} = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix} \qquad \text{and} \qquad \mathbf{A} = \begin{pmatrix} 1 & 0 \\ \rho & \sqrt{1-\rho^2} \end{pmatrix}.$$

For $-1 < \rho < 1$, write $\rho = \sin\psi$, where $\psi \in (-\pi/2, \pi/2)$. A uniform direction on the unit circle can be written as $(\cos\Theta, \sin\Theta)^{\mathsf{T}}$, where Θ is uniformly distributed on $[-\pi, \pi)$ and is independent of R. Hence

$$\boldsymbol{Y} \stackrel{\mathrm{d}}{=} R\begin{pmatrix} \cos\Theta \\ \sin\psi\cos\Theta + \cos\psi\sin\Theta \end{pmatrix} = R\begin{pmatrix} \cos\Theta \\ \sin(\Theta+\psi) \end{pmatrix}.$$

Because $R > 0$ almost surely under the stated assumption,

$$\begin{aligned} \mathbb{P}(X_1 > 0, X_2 > 0) &= \mathbb{P}(\cos\Theta > 0, \sin(\Theta+\psi) > 0) \\ &= \mathbb{P}\left(-\psi < \Theta < \frac{\pi}{2}\right) \\ &= \frac{\pi/2+\psi}{2\pi} = \frac{1}{4} + \frac{1}{2\pi}\arcsin(\rho). \end{aligned}$$

The endpoint cases follow by continuity, or directly from comonotonicity and countermonotonicity. □

No general closed formula for Spearman's rho is known for every elliptical copula. For the Gaussian family, however, the calculation in Example 12.6 gives the following special case.

Proposition 15.6 If (U_1, U_2) has a Gaussian copula with correlation parameter ρ, then

$$\rho_S(U_1, U_2) = \frac{6}{\pi}\arcsin\left(\frac{\rho}{2}\right).$$

Lecture 16 — Monday, June 27, 2016

The Gaussian formula was derived in Example 12.6. The more striking fact is that Kendall's tau has the same simple form throughout the elliptical class and does not depend on the radial distribution.

Proposition 16.1 Let $\boldsymbol{X} \sim E_2(\mathbf{0}, \mathbf{P}, \phi)$, suppose that $\mathbb{P}(\boldsymbol{X} = \mathbf{0}) = 0$, and let $\rho = \mathbf{P}_{12}$. Then

$$\tau(X_1, X_2) = \frac{2}{\pi} \arcsin(\rho). \tag{16.1}$$

Proof. Let $\boldsymbol{X}'$ be an independent copy of $\boldsymbol{X}$, and set $\boldsymbol{Y} = \boldsymbol{X} - \boldsymbol{X}'$. By the definition of Kendall's tau and the absence of ties,

$$\tau(X_1, X_2) = 2\mathbb{P}(Y_1 Y_2 > 0) - 1. \tag{16.2}$$

Central symmetry gives $\Phi_{-\boldsymbol{X}'} = \Phi_{\boldsymbol{X}}$. Hence Theorem 15.2(1) implies

$$\begin{aligned} \Phi_{\boldsymbol{Y}}(\boldsymbol{t}) &= \Phi_{\boldsymbol{X}}(\boldsymbol{t}) \Phi_{-\boldsymbol{X}'}(\boldsymbol{t}) \\ &= \phi^2(\boldsymbol{t}^\mathsf{T} \mathbf{P} \boldsymbol{t}). \end{aligned}$$

Thus $\boldsymbol{Y} \sim E_2(\mathbf{0}, \mathbf{P}, \phi^2)$. The events in which both components of $\boldsymbol{Y}$ are positive and both are negative have equal probability, so

$$\begin{aligned} \tau(X_1, X_2) &= 4\mathbb{P}(Y_1 > 0, Y_2 > 0) - 1 \\ &= 4\left(\frac{1}{4} + \frac{1}{2\pi} \arcsin(\rho)\right) - 1 \\ &= \frac{2}{\pi} \arcsin(\rho), \end{aligned}$$

where the second equality follows from Lemma 15.5. □

Remark 16.2 The assumption that the margins are continuous can be weakened. As shown by Lindskog, McNeil, and Schmock (2003), if $\boldsymbol{X} \sim E_d(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \phi)$ is componentwise nondegenerate, then for distinct i and j,

$$\tau(X_i, X_j) = \left(1 - \sum_{x \in \mathbb{R}} \mathbb{P}(X_i = x)^2\right) \frac{2}{\pi} \arcsin(\rho_{ij}). \tag{16.3}$$

The factor in parentheses is the probability that two independent copies of X_i are unequal. It equals one for continuous margins, recovering (16.1).

There is no single tail-dependence formula for all elliptical distributions: the coefficient is governed by the radial tail. Regular variation gives an important general case.

Theorem 16.3 Let $\boldsymbol{X} \sim E_d(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \phi)$. Suppose that its radial variable R is regularly varying at infinity with index $\alpha > 0$, meaning that

$$\lim_{x\to\infty} \frac{\overline{F}_R(tx)}{\overline{F}_R(x)} = t^{-\alpha} \tag{16.4}$$

for every $t > 0$, and suppose that $\mathbb{P}(R = 0) = 0$, so the margins are continuous. If $\sigma_{ii} > 0$, $\sigma_{jj} > 0$, and $|\rho_{ij}| < 1$, then

$$\lambda_L(X_i, X_j) = \lambda_U(X_i, X_j) = \frac{\displaystyle\int_{(\pi/2-\arcsin(\rho_{ij}))/2}^{\pi/2} (\cos s)^\alpha \,\mathrm{d}s}{\displaystyle\int_0^{\pi/2} (\cos s)^\alpha \,\mathrm{d}s}. \tag{16.5}$$

The equality of the upper and lower coefficients follows from radial symmetry. The explicit integral, proved by Hult and Lindskog (2002), shows that even uncorrelated components can be tail dependent when the radial distribution is sufficiently heavy-tailed.

Gaussian Copulas

Definition 16.4 The **Gaussian copula** with positive-definite correlation matrix $\mathbf{P}$ is the copula of

$$\boldsymbol{X} \sim \mathcal{N}_d(\mathbf{0}, \mathbf{P}) = E_d(\mathbf{0}, \mathbf{P}, e^{-t/2}).$$

Its density generator is

$$g(t) = (2\pi)^{-d/2} e^{-t/2}. \tag{16.6}$$

If $\boldsymbol{Z} \sim \mathcal{N}_d(\mathbf{0}, \mathbf{I}_d)$ and $\mathbf{A}\mathbf{A}^\top = \mathbf{P}$, then

$$\boldsymbol{X} = \mathbf{A}\boldsymbol{Z} = \|\boldsymbol{Z}\| \mathbf{A} \frac{\boldsymbol{Z}}{\|\boldsymbol{Z}\|}.$$

Thus the radial variable is

$$R \overset{\mathrm{d}}{=} \|\boldsymbol{Z}\| \sim \chi_d \qquad \text{and} \qquad R^2 \sim \chi^2_d. \tag{16.7}$$

The correlation matrix has $\binom{d}{2}$ free entries. To simulate from the Gaussian copula, compute a factor $\mathbf{A}$, generate d independent standard normal variables as the components of $\boldsymbol{Z}$, set $\boldsymbol{X} = \mathbf{A}\boldsymbol{Z}$, and return $(\Phi(X_1), \ldots, \Phi(X_d))$.

Proposition 16.5 For a bivariate Gaussian copula with correlation parameter $\rho \in [-1, 1]$,

$$\lambda_L = \lambda_U = \begin{cases} 0, & -1 \leq \rho < 1, \\ 1, & \rho = 1. \end{cases} \tag{16.8}$$

Proof. Radial symmetry gives $\lambda_L = \lambda_U$. For $-1 < \rho < 1$, exchangeability and Proposition 13.5(2) give

$$\lambda_L = 2 \lim_{u \downarrow 0} D_1 C(u, u) = 2 \lim_{x \to -\infty} \mathbb{P}(X_2 \leq x \mid X_1 = x).$$

The conditional normal distribution has mean ρx and variance $1 - \rho^2$, so

$$\begin{aligned} \lambda_L &= 2 \lim_{x \to -\infty} \Phi\left(\frac{x - \rho x}{\sqrt{1 - \rho^2}}\right) \\ &= 2 \lim_{x \to -\infty} \Phi\left(x \sqrt{\frac{1 - \rho}{1 + \rho}}\right) = 0. \end{aligned}$$

When $\rho = -1$, the copula is W, which also has zero lower and upper coefficients of tail dependence. When $\rho = 1$, the copula is M, and both coefficients equal one. □

Lecture 17 — Wednesday, June 29, 2016

Within the elliptical class, a diagonal correlation matrix does not generally imply independence. The Gaussian family is exceptional: independence of nondegenerate components characterizes the normal distribution among elliptical distributions.

Proposition 17.1 Let $\mathbf{P}$ be a positive-definite correlation matrix, and set $x_j = \Phi^{-1}(u_j)$. The Gaussian copula has distribution function

$$\begin{aligned} C_{\mathbf{P}}(\boldsymbol{u}) &= \Phi_{\mathbf{P}}\big(\Phi^{-1}(u_1), \ldots, \Phi^{-1}(u_d)\big) \\ &= \int_{-\infty}^{x_1} \cdots \int_{-\infty}^{x_d} \frac{1}{(2\pi)^{d/2}\sqrt{\det(\mathbf{P})}} \exp\left(-\frac{1}{2}\boldsymbol{x}^\top \mathbf{P}^{-1}\boldsymbol{x}\right) \mathrm{d}x_d \cdots \mathrm{d}x_1, \end{aligned} \tag{17.1}$$

and density

$$c_{\mathbf{P}}(\boldsymbol{u}) = \frac{1}{\sqrt{\det(\mathbf{P})}} \exp\left[-\frac{1}{2}\boldsymbol{x}^{\mathsf{T}}(\mathbf{P}^{-1} - \mathbf{I}_d)\boldsymbol{x}\right], \quad \boldsymbol{x} = \big(\Phi^{-1}(u_1), \ldots, \Phi^{-1}(u_d)\big). \tag{17.2}$$

Proof. Formula (17.1) follows directly from Theorem 5.2. Differentiating, or equivalently dividing the joint normal density by the product of its standard normal marginal densities, gives

$$c_{\mathbf{P}}(\boldsymbol{u}) = \frac{(2\pi)^{-d/2}\det(\mathbf{P})^{-1/2}\exp\big(-\boldsymbol{x}^{\mathsf{T}}\mathbf{P}^{-1}\boldsymbol{x}/2\big)}{\prod_{j=1}^{d}(2\pi)^{-1/2}\exp\Big(-x_j^2/2\Big)},$$

which simplifies to (17.2). □

For the bivariate Gaussian copula, Figure 14 shows how the sign of ρ rotates the region in which the density is largest. Both panels use the same colour scale, truncated at density value five; warmer colours represent larger values. The boundary is omitted because the density may be unbounded near two corners of the unit square.

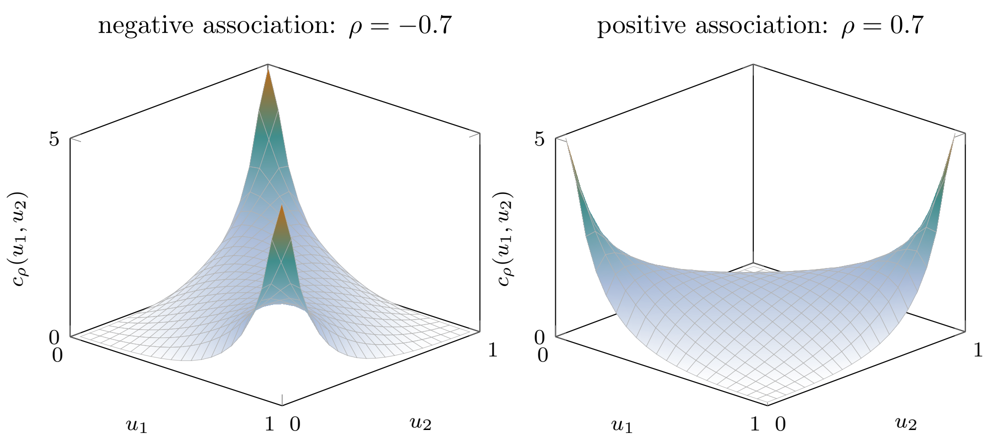


Figure 14

Reflecting either coordinate about 1/2 transforms one panel into the other, in agreement with the change from ρ to $-\rho$.

Student t Copulas

Definition 17.2 Let $\nu > 0$, let $V \sim \chi^2_\nu$, and set $W = \nu/V$. Independently, let $\boldsymbol{Z} \sim \mathcal{N}_d(\mathbf{0}, \mathbf{I}_d)$, and let $\mathbf{A}\mathbf{A}^\mathsf{T} = \mathbf{\Sigma}$. The random vector

$$\boldsymbol{X} = \boldsymbol{\mu} + \sqrt{W}\mathbf{A}\boldsymbol{Z} \tag{17.3}$$

has a multivariate Student t distribution with ν degrees of freedom, written $t_{\nu,d}(\boldsymbol{\mu}, \mathbf{\Sigma})$. When $\boldsymbol{\mu} = \mathbf{0}$ and $\mathbf{\Sigma} = \mathbf{P}$ is a correlation matrix, the copula of $\boldsymbol{X}$ is the **Student t copula** with parameters $(\nu, \mathbf{P})$.

The mixing variable W has an inverse-gamma distribution with shape and scale parameters $\nu/2$. Conditional on W, the vector in (17.3) is normal. Consequently, it is elliptical with characteristic generator and density generator

$$\phi_\nu(t) = \mathbb{E}\left[e^{-tW/2}\right], \tag{17.4}$$

$$g_{\nu,d}(t) = \frac{\Gamma((\nu+d)/2)}{(\pi\nu)^{d/2}\Gamma(\nu/2)}\left(1 + \frac{t}{\nu}\right)^{-(\nu+d)/2}. \tag{17.5}$$

Moreover,

$$R \overset{\mathrm{d}}{=} \sqrt{W}\|\boldsymbol{Z}\| \qquad \text{and} \qquad \frac{R^2}{d} = \frac{\chi^2_d/d}{\chi^2_\nu/\nu} \sim F_{d,\nu}. \tag{17.6}$$

Proposition 17.3 A d-dimensional Student t copula has $\binom{d}{2} + 1$ free parameters. The following procedure generates one observation.

1. Compute $\mathbf{A}$ such that $\mathbf{A}\mathbf{A}^\mathsf{T} = \mathbf{P}$.
2. Generate $\boldsymbol{Z} \sim \mathcal{N}_d(\mathbf{0}, \mathbf{I}_d)$.
3. Independently generate $V \sim \chi^2_\nu$.
4. Set $\boldsymbol{X} = \sqrt{\nu/V}\mathbf{A}\boldsymbol{Z}$, and return

$$\big(t_\nu(X_1), \ldots, t_\nu(X_d)\big),$$

where t_ν is the univariate Student t distribution function.

Proof. The first three steps produce the normal variance mixture in (17.3) with scale matrix $\mathbf{P}$. Since the diagonal entries of $\mathbf{P}$ equal one, each component has the univariate Student t distribution with ν degrees of freedom. The final probability transform therefore gives uniform margins and preserves the copula. $\square$

Proposition 17.4 If $\boldsymbol{X} \sim t_{\nu,d}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, then

$$f_{\boldsymbol{X}}(\boldsymbol{x}) = \frac{\Gamma((\nu+d)/2)}{\Gamma(\nu/2)(\pi\nu)^{d/2}\sqrt{\det(\boldsymbol{\Sigma})}} \left(1 + \frac{(\boldsymbol{x}-\boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})}{\nu}\right)^{-(\nu+d)/2}. \tag{17.7}$$

Proof. Let $q = (\boldsymbol{x}-\boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})$. Conditional on $V = v$, $\boldsymbol{X} \sim \mathcal{N}_d(\boldsymbol{\mu}, \nu\boldsymbol{\Sigma}/v)$, and therefore

$$\begin{aligned} f_{\boldsymbol{X}}(\boldsymbol{x}) &= \int_0^\infty f_{\boldsymbol{X}|V}(\boldsymbol{x} \mid v) f_V(v)\,\mathrm{d}v \\ &= \frac{1}{(2\pi\nu)^{d/2}\sqrt{\det(\boldsymbol{\Sigma})}2^{\nu/2}\Gamma(\nu/2)} \int_0^\infty v^{(\nu+d)/2-1} \exp\left[-\frac{v}{2}\left(1+\frac{q}{\nu}\right)\right] \mathrm{d}v. \end{aligned}$$

The integral is a gamma integral with shape $(\nu+d)/2$ and rate $(1+q/\nu)/2$. Evaluating it and simplifying the constants yields (17.7). $\square$

Stirling's formula,

$$\Gamma(x) = \sqrt{\frac{2\pi}{x}}\left(\frac{x}{e}\right)^x \left(1 + O(x^{-1})\right), \quad x \to \infty,$$

shows from (17.7) that the density converges pointwise to the $\mathcal{N}_d(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ density as $\nu \to \infty$. Accordingly, the Student t copula converges to the Gaussian copula with the same correlation matrix.

Proposition 17.5 For a bivariate Student t copula with $\nu > 0$ degrees of freedom and correlation parameter $\rho \in [-1, 1]$,

$$\lambda_L = \lambda_U = 2t_{\nu+1}\left(-\sqrt{\frac{(\nu+1)(1-\rho)}{1+\rho}}\right), \tag{17.8}$$

with the endpoint values interpreted by continuity.

Proof. Radial symmetry gives $\lambda_L = \lambda_U$. For $-1 < \rho < 1$, the same derivative argument used in

Proposition 16.5 gives

$$\lambda_L = 2 \lim_{x \to -\infty} \mathbb{P}(X_2 \le x \mid X_1 = x).$$

For a bivariate Student t vector, the conditional distribution of X_2 given $X_1 = x$ is Student t with $\nu + 1$ degrees of freedom, location ρx, and scale

$$\sqrt{1-\rho^2}\sqrt{\frac{\nu + x^2}{\nu + 1}}.$$

It follows that

$$\begin{aligned}
\lambda_L &= 2 \lim_{x \to -\infty} t_{\nu+1}\left(\frac{x(1-\rho)}{\sqrt{1-\rho^2}\sqrt{(\nu + x^2)/(\nu+1)}}\right) \\
&= 2t_{\nu+1}\left(-\sqrt{\frac{(\nu+1)(1-\rho)}{1+\rho}}\right).
\end{aligned}$$

At $\rho = -1$ the value is zero, while at $\rho = 1$ it is one. □

Proposition 17.6 Let $t_{\nu,\mathbf{P}}$ denote the distribution function of $t_{\nu,d}(\mathbf{0}, \mathbf{P})$, and set $x_j = t_\nu^{-1}(u_j)$. Then

$$\begin{aligned}
C_{\nu,\mathbf{P}}(\boldsymbol{u}) &= t_{\nu,\mathbf{P}}(x_1, \ldots, x_d) \\
&= \int_{-\infty}^{x_1} \cdots \int_{-\infty}^{x_d} \frac{\Gamma((\nu+d)/2)}{\Gamma(\nu/2)(\pi\nu)^{d/2}\sqrt{\det(\mathbf{P})}} \left(1 + \frac{\boldsymbol{x}^\mathsf{T}\mathbf{P}^{-1}\boldsymbol{x}}{\nu}\right)^{-(\nu+d)/2} \mathrm{d}x_d \cdots \mathrm{d}x_1.
\end{aligned} \tag{17.9}$$

Its density is

$$\begin{aligned}
c_{\nu,\mathbf{P}}(\boldsymbol{u}) = &\frac{\Gamma((\nu+d)/2)\Gamma(\nu/2)^{d-1}}{\Gamma((\nu+1)/2)^d\sqrt{\det(\mathbf{P})}} \\
&\times \frac{\prod_{j=1}^{d}\left(1 + \frac{x_j^2}{\nu}\right)^{(\nu+1)/2}}{\left(1 + \frac{\boldsymbol{x}^\mathsf{T}\mathbf{P}^{-1}\boldsymbol{x}}{\nu}\right)^{(\nu+d)/2}}.
\end{aligned} \tag{17.10}$$

Proof. Equation (17.9) is Sklar's factorization applied to Proposition 17.4. Dividing that joint density by the product of the d univariate Student t densities gives (17.10). □

At correlation $\rho = 0.7$, Figure 15 compares the bivariate Student t copula density for two choices of ν. The common colour scale is the same one used in Figure 14 and is truncated at density value five.

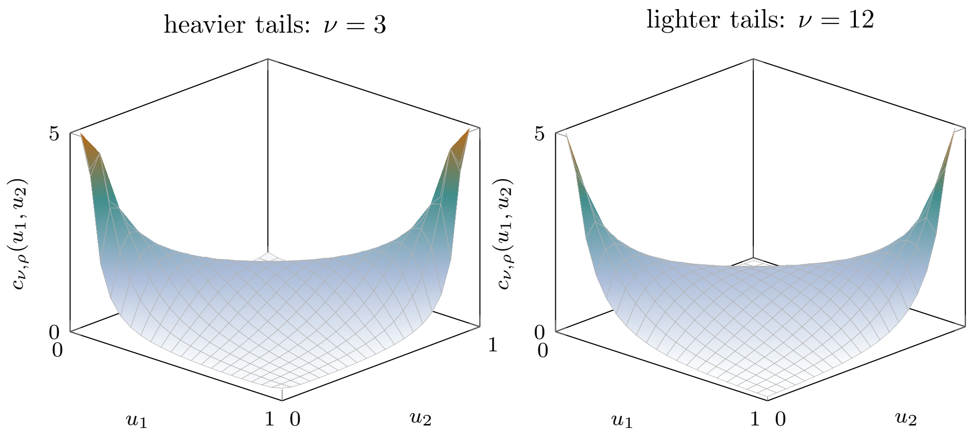


FIGURE 15

As ν increases, the density approaches the positively associated Gaussian density in Figure 14; for finite ν, substantially more density remains in the two corners where the variables enter the same tail.

Remark 17.7 The construction can use different mixing variables for different coordinate groups. Let $0 = d_0 < d_1 < \cdots < d_S = d$, let $\boldsymbol{Y} \sim \mathcal{N}_d(\mathbf{0}, \mathbf{P})$, and define

$$X_j = \sqrt{W_s} Y_j \quad \text{when } d_{s-1} < j \leq d_s. \tag{17.11}$$

The vector $(W_1, \ldots, W_S)$ is independent of $\boldsymbol{Y}$. Its components may be mutually dependent, but W_s has the inverse-gamma distribution associated with degrees of freedom ν_s. The resulting copula is a **grouped Student t copula**; when every group contains one coordinate, it is called a **generalized Student t copula**. See Demarta and McNeil (2005) for these extensions. Implementation details and common parameterization pitfalls are discussed by Hofert (2013).

Lecture 18 — Monday, July 04, 2016

5. Archimedean Copulas

Generators and Complete Monotonicity

Definition 18.1 An **Archimedean generator** is a continuous, nonincreasing function $\psi\colon [0,\infty] \to [0,1]$ such that

$$\psi(0) = 1 \qquad \text{and} \qquad \psi(\infty) = \lim_{t\to\infty} \psi(t) = 0,$$

and ψ is strictly decreasing on $[0, t_\psi]$, where

$$t_\psi = \inf\{t \geq 0 : \psi(t) = 0\}.$$

The class of these functions is denoted by Ψ. The generalized inverse is defined on $[0,1]$ by

$$\psi^{-1}(u) = \inf\{t \geq 0 : \psi(t) \leq u\},$$

so $\psi^{-1}(0) = t_\psi$. A copula is **Archimedean** if it can be written as

$$C_\psi(\boldsymbol{u}) = \psi\left(\sum_{j=1}^{d} \psi^{-1}(u_j)\right) \tag{18.1}$$

for a generator ψ having the monotonicity required in dimension d.

If ψ is strict and twice continuously differentiable, with $\psi' < 0$, differentiating (18.1) gives the bivariate density

$$c_\psi(u_1, u_2) = \frac{\psi''\big(\psi^{-1}(u_1) + \psi^{-1}(u_2)\big)}{\psi'\big(\psi^{-1}(u_1)\big)\psi'\big(\psi^{-1}(u_2)\big)}. \tag{18.2}$$

Three standard choices are the **Clayton**, **Gumbel**, and **Frank** generators, respectively:

$$\begin{aligned} \psi_\theta^{\mathrm{Cl}}(t) &= (1+t)^{-1/\theta}, && \theta > 0, \\ \psi_\theta^{\mathrm{Gu}}(t) &= \exp\big(-t^{1/\theta}\big), && \theta \geq 1, \end{aligned}$$

$$\psi_\theta^{\mathrm{Fr}}(t) = -\frac{1}{\theta}\log\bigl(1-(1-e^{-\theta})e^{-t}\bigr), \qquad \theta > 0.$$

The corresponding densities are compared in Figure 16 at representative positive dependence parameters. The panels share the same colour scale, truncated at density value five; warmer colours represent larger values. Together, the panels make Clayton's lower-tail concentration, Gumbel's upper-tail concentration, and Frank's more diffuse dependence immediately visible.

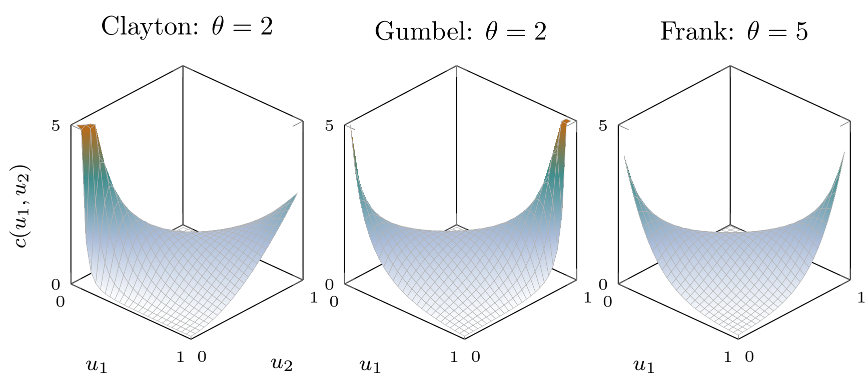


Figure 16

This contrast anticipates the general tail formulas in Proposition 19.3.

Remark 18.2 The following observations explain several basic features of the class.

1. For every $c > 0$, the rescaled generator $\psi_c(t) = \psi(ct)$ generates the same copula as ψ, because $\psi_c^{-1}(u) = \psi^{-1}(u)/c$.
2. Define the serial iterates by

$$\begin{aligned} C^2(u_1,u_2) &= \psi\bigl(\psi^{-1}(u_1)+\psi^{-1}(u_2)\bigr), \\ C^{n+1}(u_1,\dots,u_{n+1}) &= C^2\bigl(C^n(u_1,\dots,u_n),u_{n+1}\bigr), \quad n \ge 2. \end{aligned} \tag{18.3}$$

If $u, v \in (0,1)$, we can choose a positive integer n such that $n\psi^{-1}(u) > \psi^{-1}(v)$. It follows that

$$C^n(u,\dots,u) = \psi\bigl(n\psi^{-1}(u)\bigr) < v. \tag{18.4}$$

This analogue of the Archimedean property for multiplication motivates the name of the copula class.

3. The generator $\psi(t) = e^{-t}$ produces the independence copula in every dimension, since

$$e^{-\sum_{j=1}^{d}(-\log u_j)} = \prod_{j=1}^{d} u_j.$$

4. The upper Fréchet–Hoeffding bound is not Archimedean. Indeed, for $u \in (0,1)$ and $d \geq 2$, strict decrease of the generator gives

$$C_\psi(u, \ldots, u) = \psi\big(d\psi^{-1}(u)\big) < u,$$

whereas $M(u, \ldots, u) = u$.

Definition 18.3 Let $a < b$ be extended real numbers, and let $f\colon [a,b] \to \mathbb{R}$.

1. The function f is **absolutely monotone** if it is continuous on $[a,b]$, infinitely differentiable on (a,b), and $f^{(k)}(x) \geq 0$ for every nonnegative integer k and every $x \in (a,b)$.
2. For an integer $d \geq 2$, the function f is **d-monotone** if it is continuous on $[a,b]$, is $(d-2)$-times differentiable on (a,b), satisfies

$$(-1)^k f^{(k)}(x) \geq 0, \quad k \in \{0, \ldots, d-2\},$$

and $(-1)^{d-2} f^{(d-2)}$ is nonincreasing and convex.
3. The function f is **completely monotone** if it is d-monotone for every integer $d \geq 2$. Equivalently, it is infinitely differentiable on (a,b) and $(-1)^k f^{(k)} \geq 0$ for every nonnegative integer k.

We write

$$\Psi_d = \{\psi \in \Psi : \psi \text{ is } d\text{-monotone}\}, \tag{18.5}$$
$$\Psi_\infty = \{\psi \in \Psi : \psi \text{ is completely monotone}\}. \tag{18.6}$$

Theorem 18.4 (Bernstein–Widder theorem) A generator ψ belongs to Ψ_∞ if and only if it is the Laplace transform of a unique probability distribution F on $[0,\infty)$; that is,

$$\psi(t) = \mathscr{L}[F](t) = \int_{[0,\infty)} e^{-tv}\,\mathrm{d}F(v) = \mathbb{E}[e^{-tV}], \quad V \sim F. \tag{18.7}$$

This classical transform characterization is the source of the frailty representation below. A modern copula treatment appears in McNeil and Nešlehová (2009).

Corollary 18.5 Let $\psi \in \Psi_\infty$, and let $F = \mathscr{L}^{-1}[\psi]$ be the distribution in Theorem 18.4. Then the following statements hold.

1. The distribution has no atom at zero: $F(0) = 0$.
2. The generator is strict: $\psi(t) > 0$ for every finite $t \geq 0$.

Proof. By dominated convergence,

$$\lim_{t\to\infty} \psi(t) = \mathbb{P}(V = 0) = F(0).$$

The defining condition $\psi(\infty) = 0$ therefore proves Part (1). For Part (2), choose $v_0 > 0$ such that $F(v_0) > 0$. For each finite $t \geq 0$,

$$\psi(t) \geq \mathbb{E}\left[e^{-tV}\mathbb{1}_{\{V\leq v_0\}}\right] \geq e^{-tv_0}F(v_0) > 0.$$

□

Theorem 18.6 Let $\psi \in \Psi$. The function

$$C_d(\boldsymbol{u}) = \psi\left(\sum_{j=1}^{d} \psi^{-1}(u_j)\right) \tag{18.8}$$

is a copula for every integer $d \geq 2$ if and only if $\psi \in \Psi_\infty$.

Proof. *Forward implication.* Suppose that C_d is a copula in every dimension. A finite difference criterion for complete monotonicity states that ψ is completely monotone if

$$\sum_{k=0}^{n} \binom{n}{k} (-1)^{n-k} \psi(x - kh) \geq 0 \tag{18.9}$$

whenever n is a positive integer and $x - nh \geq 0$. To obtain this inequality from a copula volume, put

$$s = \frac{x - nh}{n}, \qquad a = \psi(s + h), \qquad \text{and} \qquad b = \psi(s).$$

At a vertex of the box $(a, b]^n$ having k upper coordinates and $n - k$ lower coordinates, the sum of the inverse generator arguments is

$$ks + (n - k)(s + h) = x - kh.$$

Therefore,

$$\Delta_{((a,\dots,a),(b,\dots,b)]} C_n = \sum_{k=0}^{n} \binom{n}{k} (-1)^{n-k} \psi(x - kh) \geq 0.$$

The finite difference criterion proves that ψ is completely monotone.

Reverse implication. Suppose that $\psi \in \Psi_\infty$. By Theorem 18.4, there is a distribution F on $(0, \infty)$ such that

$$\begin{aligned} C_d(\boldsymbol{u}) &= \int_0^\infty \exp\left[-v \sum_{j=1}^{d} \psi^{-1}(u_j) \right] \mathrm{d}F(v) \\ &= \int_0^\infty \prod_{j=1}^{d} G_v(u_j) \, \mathrm{d}F(v), \end{aligned} \tag{18.10}$$

where

$$G_v(u) = \exp\bigl(-v\psi^{-1}(u)\bigr).$$

For each $v > 0$, the function G_v is a distribution function on $[0, 1]$, so the integrand in (18.10) is a product distribution and its mixture is a multivariate distribution function. Its jth margin is

$$\int_0^\infty G_v(u) \, \mathrm{d}F(v) = \psi(\psi^{-1}(u)) = u.$$

Thus C_d is a copula in every dimension. □

The Laplace transform representation also gives a direct sampling method. The common random variable V is called a **frailty**; it creates dependence by acting simultaneously on all coordinates.

Proposition 18.7 Let $\psi \in \Psi_\infty$, and let F be the mixing distribution in (18.7). A random vector with the Archimedean copula generated by ψ can be simulated as follows.

1. Generate $V \sim F$.
2. Independently of V, generate independent random variables $E_1, \dots, E_d$, each having the standard exponential distribution.
3. Return
$$\boldsymbol{U} = \left(\psi\left(\frac{E_1}{V}\right), \dots, \psi\left(\frac{E_d}{V}\right)\right). \tag{18.11}$$

The construction with shared frailty in Proposition 18.7 is illustrated in Figure 17.

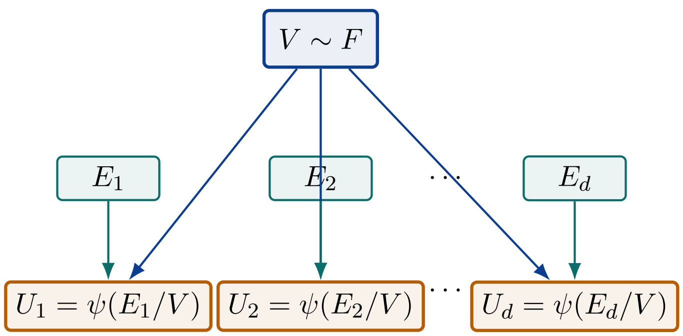


FIGURE 17

Proof. Because ψ is decreasing, conditioning on $V = v$ gives

$$\begin{aligned}\mathbb{P}(\boldsymbol{U} \le \boldsymbol{u} \mid V = v) &= \mathbb{P}\left(E_j \ge v\psi^{-1}(u_j) \text{ for } j = 1, \dots, d\right) \\ &= \prod_{j=1}^d \exp\left(-v\psi^{-1}(u_j)\right) \\ &= \exp\left(-v\sum_{j=1}^d \psi^{-1}(u_j)\right).\end{aligned}$$

Integrating with respect to F and applying (18.7) yields

$$\mathbb{P}(\boldsymbol{U} \le \boldsymbol{u}) = \int_0^\infty \exp\left(-v\sum_{j=1}^d \psi^{-1}(u_j)\right) \mathrm{d}F(v) = \psi\left(\sum_{j=1}^d \psi^{-1}(u_j)\right).$$

This is precisely the copula in (18.1). □

Lemma 18.8 For $k \in \{0, 1\}$, let C_k be a d-dimensional Archimedean copula with strict generator ψ_k. Then

$$C_0 \le C_1 \iff \psi_0^{-1} \circ \psi_1 \text{ is subadditive on } [0, \infty). \tag{18.12}$$

Proof. Since ψ_0 is decreasing, the pointwise inequality $C_0(\boldsymbol{u}) \le C_1(\boldsymbol{u})$ is equivalent to

$$\sum_{j=1}^{d} \psi_0^{-1}(u_j) \ge \psi_0^{-1}\left(\psi_1\left(\sum_{j=1}^{d} \psi_1^{-1}(u_j)\right)\right).$$

Set $t_j = \psi_1^{-1}(u_j)$ and $f = \psi_0^{-1} \circ \psi_1$. Because ψ_1 is strict, the variables $t_1, \ldots, t_d$ range independently over $[0, \infty)$. The preceding inequality therefore becomes

$$f\left(\sum_{j=1}^{d} t_j\right) \le \sum_{j=1}^{d} f(t_j), \quad t_1, \ldots, t_d \ge 0.$$

This is finite subadditivity of f, which is equivalent by induction to ordinary subadditivity for two terms. □

Lecture 19 — Wednesday, July 06, 2016

Concordance and Tail Dependence

Complete monotonicity forces an Archimedean copula to lie above the independence copula. Consequently, every concordance measure that is monotone under the pointwise copula order is nonnegative on this class.

Proposition 19.1 Let C be an Archimedean copula with generator $\psi \in \Psi_\infty$. Then

$$C(\boldsymbol{u}) \ge \Pi(\boldsymbol{u}) \quad \text{for every } \boldsymbol{u} \in [0, 1]^d. \tag{19.1}$$

Proof. Write $\psi(t) = \int_0^\infty e^{-tv}\,\mathrm{d}F(v)$, as in Theorem 18.4. For $t > 0$ and arbitrary $a, b \in \mathbb{R}$, differentiation under the integral is justified by the exponential factor, and

$$\begin{pmatrix} a & b \end{pmatrix} \begin{pmatrix} \psi(t) & \psi'(t) \\ \psi'(t) & \psi''(t) \end{pmatrix} \begin{pmatrix} a \\ b \end{pmatrix} = \int_0^\infty (a - bv)^2 e^{-tv}\,\mathrm{d}F(v) \geq 0.$$

Thus the displayed matrix is positive semidefinite, and hence

$$\psi(t)\psi''(t) - \psi'(t)^2 \geq 0.$$

Strictness of ψ, proved in Corollary 18.5(2), now gives

$$(\log \psi)''(t) = \frac{\psi(t)\psi''(t) - \psi'(t)^2}{\psi(t)^2} \geq 0.$$

Therefore $f = -\log \psi$ is concave on $(0, \infty)$, and continuity at zero extends the concavity to $[0, \infty)$, with $f(0) = 0$.

If $x, y \geq 0$ and $x + y > 0$, concavity gives

$$f(x) \geq \frac{x}{x+y} f(x+y) \qquad \text{and} \qquad f(y) \geq \frac{y}{x+y} f(x+y).$$

Adding these inequalities proves that f is subadditive. The independence generator is $\psi_0(t) = e^{-t}$, and $\psi_0^{-1} \circ \psi = -\log \psi = f$. The conclusion now follows from Lemma 18.8. □

Unlike Kendall's tau, Spearman's rho does not generally reduce to an integral in one dimension involving only the generator. For a smooth strict generator, Kendall's tau has the following two useful forms.

Proposition 19.2 Let C be a bivariate Archimedean copula with a strict, twice continuously differentiable generator ψ. Then

$$\tau(C) = 1 - 4\int_0^\infty t\psi'(t)^2\,\mathrm{d}t = 1 + 4\int_0^1 \frac{\psi^{-1}(s)}{(\psi^{-1})'(s)}\,\mathrm{d}s. \tag{19.2}$$

Proof. By Proposition 12.5(2),

$$\tau(C) = 4\int_{[0,1]^2} C(\boldsymbol{u})\,\mathrm{d}C(\boldsymbol{u}) - 1.$$

Put $x_j = \psi^{-1}(u_j)$. The density factors involving the derivatives of ψ^{-1} cancel with the Jacobian, so

$$\int_{[0,1]^2} C(\boldsymbol{u})\,\mathrm{d}C(\boldsymbol{u}) = \int_0^\infty \int_0^\infty \psi(x_1+x_2)\psi''(x_1+x_2)\,\mathrm{d}x_1\,\mathrm{d}x_2 = \int_0^\infty t\psi(t)\psi''(t)\,\mathrm{d}t.$$

Integration by parts, together with $\psi(0)=1$, $\psi(\infty)=0$, and the corresponding boundary limits, yields

$$\int_0^\infty t\psi(t)\psi''(t)\,\mathrm{d}t = -\int_0^\infty \psi(t)\psi'(t)\,\mathrm{d}t - \int_0^\infty t\psi'(t)^2\,\mathrm{d}t = \frac{1}{2} - \int_0^\infty t\psi'(t)^2\,\mathrm{d}t.$$

This proves the first equality in (19.2). Finally, the substitution $s = \psi(t)$, or equivalently $t = \psi^{-1}(s)$, gives

$$\int_0^\infty t\psi'(t)^2\,\mathrm{d}t = -\int_0^1 \frac{\psi^{-1}(s)}{(\psi^{-1})'(s)}\,\mathrm{d}s,$$

which proves the second equality. □

Proposition 19.3 Let C be a bivariate Archimedean copula with a strict generator ψ. Whenever the indicated limits exist,

$$\lambda_L = \lim_{t\to\infty} \frac{\psi(2t)}{\psi(t)}, \tag{19.3}$$

$$\lambda_U = 2 - \lim_{t\downarrow 0} \frac{1-\psi(2t)}{1-\psi(t)}. \tag{19.4}$$

Proof. The diagonal of C is

$$\delta_C(u) = C(u,u) = \psi\big(2\psi^{-1}(u)\big).$$

In the lower tail, set $u = \psi(t)$. Then $u \downarrow 0$ if and only if $t \to \infty$, and Proposition 13.5(1) gives

$$\lambda_L = \lim_{u\downarrow 0} \frac{\delta_C(u)}{u} = \lim_{t\to\infty} \frac{\psi(2t)}{\psi(t)}.$$

Similarly, $u \uparrow 1$ if and only if $t \downarrow 0$, so Proposition 13.5(3) gives

$$\lambda_U = 2 - \lim_{u\uparrow 1} \frac{1-\delta_C(u)}{1-u} = 2 - \lim_{t\downarrow 0} \frac{1-\psi(2t)}{1-\psi(t)}.$$

□

Proposition 19.4 Let $\psi \in \Psi_\infty$.

1. Suppose that
$$\psi(t) = \sum_{k=1}^{\infty} p_k e^{-x_k t} \qquad \text{and} \qquad 0 < x_1 < x_2 < \cdots,$$
where $(p_k)_{k\in\mathbb{N}}$ is a probability mass function. Then $\lambda_L = 0$.
2. Suppose that the mixing variable V in (18.7) satisfies $\mathbb{E}[V] < \infty$. Then $\lambda_U = 0$.

Proof. Under the assumption in Part (1),
$$\psi(2t) = \sum_{k=1}^{\infty} p_k e^{-2x_k t} \le e^{-x_1 t} \sum_{k=1}^{\infty} p_k e^{-x_k t} = e^{-x_1 t}\psi(t).$$
Consequently,
$$0 \le \frac{\psi(2t)}{\psi(t)} \le e^{-x_1 t} \longrightarrow 0 \quad \text{as } t \to \infty.$$
Equation (19.3) proves $\lambda_L = 0$.

For Part (2), differentiation of the Laplace transform at zero gives
$$-\psi'(0+) = \int_0^\infty v \,\mathrm{d}F(v) = \mathbb{E}[V] \in (0,\infty).$$
Therefore,
$$\lim_{t\downarrow 0} \frac{1-\psi(2t)}{1-\psi(t)} = 2 \lim_{t\downarrow 0} \frac{[1-\psi(2t)]/(2t)}{[1-\psi(t)]/t} = 2.$$
Equation (19.4) now gives $\lambda_U = 0$. □

Lecture 20 — Monday, July 11, 2016

Finite-Dimensional Generators and Extensions

Parametric generator families and their transformations provide flexible models beyond the standard one-parameter examples. Many such families are available in the `copula` package for R. Two transformations have particularly transparent frailty representations.

Proposition 20.1 Let $\psi \in \Psi_\infty$, and let V have the mixing distribution $\mathscr{L}^{-1}[\psi]$.

1. For a positive integer m, the function
$$\widetilde{\psi}_m(t) = \psi(t)^m \tag{20.1}$$
belongs to Ψ_∞. Its mixing variable has the representation
$$\widetilde{V}_m \overset{\mathrm{d}}{=} \sum_{j=1}^{m} V_j,$$
where $V_1, \ldots, V_m$ are independent copies of V.
2. For $h \geq 0$ and $\alpha \in (0, 1]$, the tilted outer-power transform
$$\widetilde{\psi}(t) = \psi\big((h+t)^\alpha - h^\alpha\big) \tag{20.2}$$
also belongs to Ψ_∞. If S_α is positive α-stable, normalized by
$$\mathbb{E}[e^{-tS_\alpha}] = e^{-t^\alpha},$$
define, conditionally on $V = v$, an exponentially tilted version $\widetilde{S}$ by
$$\frac{\mathrm{d}\mathbb{P}_{\widetilde{S}|V=v}}{\mathrm{d}\mathbb{P}_{S_\alpha}}(s) = \exp\big(-hv^{1/\alpha}s + vh^\alpha\big). \tag{20.3}$$
Then a mixing variable for $\widetilde{\psi}$ is
$$\widetilde{V} = V^{1/\alpha}\widetilde{S}. \tag{20.4}$$

Proof. For Part (1), independence gives

$$\mathbb{E}\left[\exp\left(-t\sum_{j=1}^{m} V_j\right)\right] = \prod_{j=1}^{m}\mathbb{E}[e^{-tV_j}] = \psi(t)^m.$$

For Part (2), the right-hand side of (20.3) integrates to one because

$$\mathbb{E}\left[e^{-hv^{1/\alpha}S_\alpha}\right] = e^{-vh^\alpha}.$$

For $t \geq 0$, exponential tilting then gives

$$\mathbb{E}\left[e^{-tV^{1/\alpha}\widetilde{S}} \mid V = v\right] = \exp(vh^\alpha)\mathbb{E}\left[e^{-(h+t)v^{1/\alpha}S_\alpha}\right] = \exp\bigl(-v((h+t)^\alpha - h^\alpha)\bigr).$$

Taking expectation with respect to V produces (20.2). The special case $\alpha = 1$ is interpreted using $S_1 = 1$ almost surely. Further sampling details are given by Hofert (2011). $\square$

Exercise 20.2 Let $\psi \in \Psi_\infty$, let $\alpha \in (0,1]$, and define $\widetilde{\psi}(t) = \psi(t^\alpha)$.

1. Show that the Kendall's tau values of the original and transformed copulas satisfy

$$\widetilde{\tau} = 1 - \alpha(1-\tau).$$

2. For the Clayton generator $\psi(t) = (1+t)^{-1/\theta}$, where $\theta > 0$, show that

$$\widetilde{\lambda}_L = 2^{-\alpha/\theta} \qquad \text{and} \qquad \widetilde{\lambda}_U = 2 - 2^\alpha.$$

Complete monotonicity is sufficient in every dimension, but it is stronger than necessary when the dimension is fixed. The precise finite-dimensional condition has a useful geometric interpretation.

Theorem 20.3 Let $d \geq 2$ and $\psi \in \Psi$.

1. The function in (18.1) is a d-dimensional copula if and only if $\psi \in \Psi_d$.
2. Suppose that C_ψ is a d-dimensional Archimedean copula and $\boldsymbol{U} \sim C_\psi$. Then

$$\boldsymbol{X} = \bigl(\psi^{-1}(U_1), \ldots, \psi^{-1}(U_d)\bigr) \overset{\mathrm{d}}{=} R\boldsymbol{S}, \tag{20.5}$$

where $R > 0$, $F_R(0) = 0$, and R is independent of a random vector $\boldsymbol{S}$ uniformly distributed on the unit simplex

$$\mathcal{S}_d = \{\boldsymbol{s} \in \mathbb{R}^d_+ : \|\boldsymbol{s}\|_1 = 1\}. \tag{20.6}$$

Thus $\boldsymbol{X}$ has a simplex, or ℓ_1-norm symmetric, distribution.

3. Conversely, let $\boldsymbol{X} = R\boldsymbol{S}$ have the representation in (20.5). Its survival copula is Archimedean with generator

$$\psi(t) = \int_t^\infty \left(1 - \frac{t}{r}\right)^{d-1} \mathrm{d}F_R(r) = \mathbb{E}\left[\left(1 - \frac{t}{R}\right)_+^{d-1}\right] = W_d[F_R](t). \tag{20.7}$$

This is the **Williamson d-transform** of F_R. Moreover,

$$R = \|\boldsymbol{X}\|_1 \qquad \text{and} \qquad \boldsymbol{S} = \frac{\boldsymbol{X}}{\|\boldsymbol{X}\|_1} \quad \text{almost surely,}$$

and these two quantities are independent.

Proof. For the necessity in Part (1), apply the d-increasing property of C_ψ to boxes whose endpoints are written in coordinates of the inverse generator. The resulting nonnegative divided differences of ψ are equivalent to d-monotonicity. For sufficiency, Williamson's inversion theorem states that $\psi \in \Psi_d$ if and only if there is a unique distribution F_R on $(0, \infty)$ for which (20.7) holds. This analytic result, combined with the following probabilistic argument, completes Part (1). Full details of the divided difference equivalence and the inversion theorem appear in McNeil and Nešlehová (2009).

For fixed $r > 0$, ratios of simplex volumes give

$$\mathbb{P}(rS_j > x_j \text{ for } j = 1, \ldots, d) = \left(1 - \frac{x_1 + \cdots + x_d}{r}\right)_+^{d-1}.$$

Conditioning on R therefore yields

$$\mathbb{P}(X_j > x_j \text{ for } j = 1, \ldots, d) = \psi(x_1 + \cdots + x_d). \tag{20.8}$$

In particular, every marginal survival function is ψ, and the survival copula is

$$C_\psi(u_1, \ldots, u_d) = \psi\left(\sum_{j=1}^d \psi^{-1}(u_j)\right).$$

This proves Part (3).

Conversely, if $\boldsymbol{U} \sim C_\psi$ and $X_j = \psi^{-1}(U_j)$, then monotonicity of ψ gives

$$\mathbb{P}(X_j > x_j \text{ for } j = 1, \ldots, d) = C_\psi\big(\psi(x_1), \ldots, \psi(x_d)\big) = \psi(x_1 + \cdots + x_d).$$

The uniqueness in Williamson's inversion theorem identifies this law with that of $R\boldsymbol{S}$, proving Part (2). Finally, $\|\boldsymbol{S}\|_1 = 1$ gives the identities for R and $\boldsymbol{S}$ that hold almost surely, including their asserted independence. □

For $d = 2$, the geometry behind (20.5) is shown in Figure 18. The direction $\boldsymbol{S}$ lies on the unit ℓ_1-simplex, while R controls the distance along its ray.

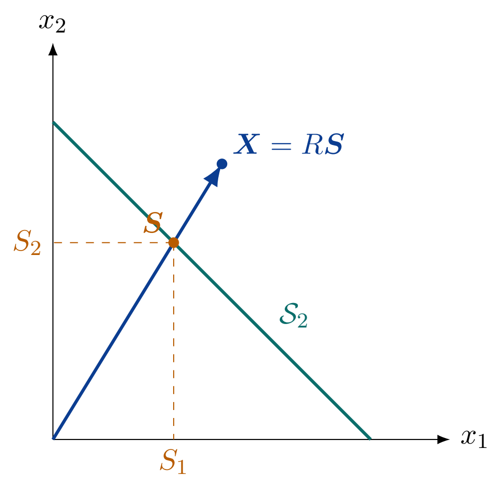


FIGURE 18

Example 20.4 For $d \geq 2$, consider

$$\psi_d(t) = (1 - t)_+^{d-1}.$$

This generator belongs to $\Psi_d \setminus \Psi_{d+1}$, so it generates copulas up to dimension d, but not in dimension $d + 1$.

Solution. If $R = 1$ almost surely, then (20.7) becomes

$$W_d[F_R](t) = (1 - t)_+^{d-1} = \psi_d(t).$$

Thus $\psi_d \in \Psi_d$. The function fails to be $(d-1)$-times differentiable at $t = 1$, so it does not belong to Ψ_{d+1}. The corresponding random vector is simply $\boldsymbol{X} = \boldsymbol{S}$. In particular, if $E_1, \ldots, E_d$ are independent standard exponential random variables, then

$$\boldsymbol{S} \overset{\mathrm{d}}{=} \left(\frac{E_1}{\sum_{j=1}^d E_j}, \ldots, \frac{E_d}{\sum_{j=1}^d E_j} \right),$$

and its survival copula is generated by ψ_d. □

Proposition 20.5 Let $\psi \in \Psi_d$. A random vector with copula C_ψ can be generated as follows.

1. Generate R from the Williamson inverse

$$F_R(x) = 1 - \sum_{k=0}^{d-2} \frac{(-x)^k \psi^{(k)}(x)}{k!} - \frac{(-x)^{d-1}\psi_+^{(d-1)}(x)}{(d-1)!}, \quad x > 0, \tag{20.9}$$

   where $\psi_+^{(d-1)}$ denotes the right-hand derivative of $\psi^{(d-2)}$.

2. Independently generate $\boldsymbol{S}$ uniformly on $\mathcal{S}_d$. We can generate independent standard exponential random variables $E_1, \ldots, E_d$ and set

$$S_j = \frac{E_j}{\sum_{k=1}^d E_k}, \quad j \in \{1, \ldots, d\}. \tag{20.10}$$

3. Return

$$U_j = \psi(RS_j) = \psi\left(R \frac{E_j}{\sum_{k=1}^d E_k} \right), \quad j \in \{1, \ldots, d\}. \tag{20.11}$$

Proof. The inversion formula (20.9) is the unique inverse of the Williamson transform in (20.7). Normalizing independent gamma variables with common unit scale gives a Dirichlet random vector; standard exponentials are gamma variables with shape one, so (20.10) is uniform on $\mathcal{S}_d$. Theorem 20.3(3) then shows that $\boldsymbol{X} = R\boldsymbol{S}$ has marginal survival function ψ and survival copula C_ψ. Hence $\boldsymbol{U} = (\psi(X_1), \ldots, \psi(X_d))$ has uniform margins and copula C_ψ, as asserted. □

Although exchangeability makes Archimedean copulas tractable, it also forces all pairwise marginal copulas to be identical. Two important extensions relax this symmetry.

Remark 20.6 Replace the uniform simplex direction in (20.5) by

$$\boldsymbol{S} = \left(\frac{Z_1}{\sum_{k=1}^d Z_k}, \dots, \frac{Z_d}{\sum_{k=1}^d Z_k} \right),$$

where $Z_1, \dots, Z_d$ are independent gamma random variables with common unit scale and respective shapes $\alpha_1, \dots, \alpha_d > 0$. Then $\boldsymbol{S}$ has the Dirichlet distribution with parameter vector $(\alpha_1, \dots, \alpha_d)$. The survival copula of $R\boldsymbol{S}$ is called a **Liouville copula**. When all shapes equal one, this construction reduces to the Archimedean case; unequal shapes generally produce a nonexchangeable copula. See McNeil and Nešlehová (2010).

Nested Archimedean copulas instead combine several Archimedean building blocks. Given disjoint coordinate groups $\boldsymbol{u}_1, \dots, \boldsymbol{u}_k$, we consider

$$C(\boldsymbol{u}) = C_0\big(C_1(\boldsymbol{u}_1), \dots, C_k(\boldsymbol{u}_k)\big). \tag{20.12}$$

The expression in (20.12) is not automatically a copula; compatibility conditions between parent and child generators are required. The construction originated with Joe (1997), and its stochastic representation was developed by McNeil (2008).

Proposition 20.7 Let $\psi_0, \psi_1 \in \Psi_\infty$, and define the three-dimensional nested Archimedean candidate

$$C(u_1, u_2, u_3) = C_0\big(u_1, C_1(u_2, u_3)\big). \tag{20.13}$$

If

$$\big(\psi_0^{-1} \circ \psi_1\big)' \quad \text{is completely monotone,} \tag{20.14}$$

then C is a copula.

Proof. Let $F_0 = \mathscr{L}^{-1}[\psi_0]$, and, for $x > 0$, define

$$g_0(u; x) = \exp\big(-x\psi_0^{-1}(u)\big), \tag{20.15}$$
$$\psi_{0,1}(t; x) = \exp\left[-x\psi_0^{-1}\big(\psi_1(t)\big)\right]. \tag{20.16}$$

Set $f = \psi_0^{-1} \circ \psi_1$. Condition (20.14) says that f is a Bernstein function. Because $t \mapsto e^{-xt}$ is completely monotone, its composition with f is completely monotone. Thus $\psi_{0,1}(\,\cdot\,; x)$ is an Archimedean generator for every $x > 0$. Denote its copula by $C_{0,1}^{(x)}$.

Using the mixing representation of ψ_0, the expression in (20.13) becomes

$$\begin{aligned} C(u_1, u_2, u_3) &= \int_0^\infty g_0(u_1; x)\psi_{0,1}\big(\psi_1^{-1}(u_2) + \psi_1^{-1}(u_3); x\big)\,\mathrm{d}F_0(x) \\ &= \int_0^\infty g_0(u_1; x)C_{0,1}^{(x)}\big(g_0(u_2; x), g_0(u_3; x)\big)\,\mathrm{d}F_0(x). \end{aligned} \tag{20.17}$$

For the second equality, observe that

$$\psi_{0,1}^{-1}\big(g_0(u; x); x\big) = \psi_1^{-1}(u).$$

For each fixed x, the integrand in (20.17) is a three-dimensional distribution function with all three margins equal to $g_0(\,\cdot\,; x)$. Mixing over F_0 gives margins

$$\int_0^\infty g_0(u; x)\,\mathrm{d}F_0(x) = \psi_0\big(\psi_0^{-1}(u)\big) = u.$$

Hence the mixture is a copula. □

Lecture 21 — Monday, July 18, 2016

The sufficient nesting condition in (20.14) also leads to a recursive frailty representation for deeper trees. The following example contains three nesting levels.

Proposition 21.1 Consider the nested copula

$$C(\boldsymbol{u}) = C_0\left(u_1,\, C_1(u_2, u_3),\, C_2\big(u_4, C_3(u_5, u_6, u_7)\big)\right), \tag{21.1}$$

whose tree is shown in Figure 19. Suppose that the sufficient nesting condition holds along each parent–child edge. Define

$$\psi_{ij}(t; v) = \exp\left[-v\psi_i^{-1}\big(\psi_j(t)\big)\right]$$

for $(i, j) \in \{(0, 1), (0, 2), (2, 3)\}$. The following procedure generates a random vector with copula C.

1. Generate
$$V_0 \sim \mathscr{L}^{-1}[\psi_0].$$

2. Conditionally on V_0, generate V_{01} and V_{02} independently according to

$$V_{0j} \mid V_0 \sim \mathscr{L}^{-1}[\psi_{0j}(\,\cdot\,; V_0)], \quad j \in \{1, 2\}.$$

   Conditionally on V_{02}, generate

$$V_{23} \mid V_{02} \sim \mathscr{L}^{-1}[\psi_{23}(\,\cdot\,; V_{02})].$$

3. Independently of these frailties, generate independent standard exponential random variables $E_1, \dots, E_7$, and return

$$\begin{aligned} \boldsymbol{U} = \big(&\psi_0(E_1/V_0),\ \psi_1(E_2/V_{01}),\ \psi_1(E_3/V_{01}),\ \psi_2(E_4/V_{02}), \\ &\psi_3(E_5/V_{23}),\ \psi_3(E_6/V_{23}),\ \psi_3(E_7/V_{23})\big). \end{aligned} \tag{21.2}$$

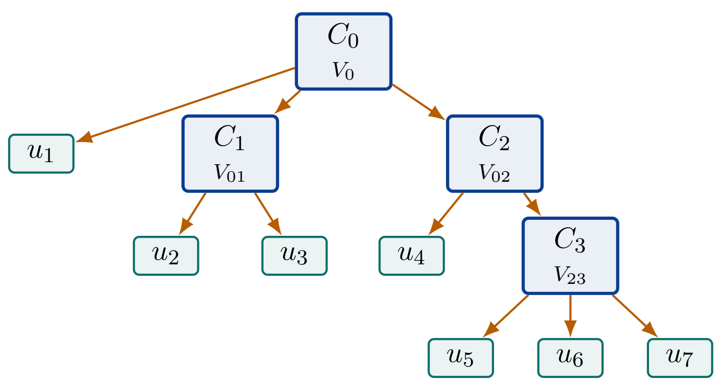


Figure 19

Proof. Condition on all four frailties. By independence of the exponential random variables, the conditional joint distribution is the product of the corresponding exponential survival probabilities. Integrating first over V_{23} reconstructs the deepest child:

$$\mathbb{E}\left[\exp\left(-V_{23}\sum_{j=5}^{7}\psi_3^{-1}(u_j)\right)\middle| V_{02}\right] = \psi_{23}\left(\sum_{j=5}^{7}\psi_3^{-1}(u_j); V_{02}\right) = \exp\left[-V_{02}\psi_2^{-1}\big(C_3(u_5, u_6, u_7)\big)\right].$$

After multiplication by the contribution from u_4, integration over V_{02} yields

$$\exp\left[-V_0\psi_0^{-1}\big(C_2(u_4, C_3(u_5, u_6, u_7))\big)\right].$$

Likewise, integrating over V_{01} turns the contributions from u_2, u_3 into

$$\exp\left[-V_0\psi_0^{-1}\big(C_1(u_2,u_3)\big)\right].$$

Including the leaf u_1 and finally integrating over V_0 gives

$$\psi_0\left(\psi_0^{-1}(u_1)+\psi_0^{-1}\big(C_1(u_2,u_3)\big)+\psi_0^{-1}\big(C_2(u_4,C_3(u_5,u_6,u_7))\big)\right),$$

which is exactly (21.1). □

Remark 21.2 The conditional mixing laws deeper in the tree, such as the law of $V_{23} \mid V_{02}$, often have no elementary density or quantile function. Practical simulation may therefore require specialized methods for exponentially tilted stable laws or numerical Laplace inversion; see Hofert (2011).

Solution to Exercise 20.2. For Part (1), differentiation gives

$$\widetilde{\psi}'(t)=\alpha t^{\alpha-1}\psi'(t^\alpha).$$

Using (19.2) and then substituting $s=t^\alpha$, we obtain

$$\begin{aligned}1-\widetilde{\tau}&=4\int_0^\infty t\widetilde{\psi}'(t)^2\,\mathrm{d}t\\&=4\alpha^2\int_0^\infty t^{2\alpha-1}\psi'(t^\alpha)^2\,\mathrm{d}t\\&=4\alpha\int_0^\infty s\psi'(s)^2\,\mathrm{d}s=\alpha(1-\tau).\end{aligned}$$

Hence

$$\widetilde{\tau}=1-\alpha(1-\tau).$$

For Part (2),

$$\widetilde{\psi}(t)=(1+t^\alpha)^{-1/\theta}.$$

Equation (19.3) gives

$$\widetilde{\lambda}_L=\lim_{t\to\infty}\left(\frac{1+2^\alpha t^\alpha}{1+t^\alpha}\right)^{-1/\theta}=2^{-\alpha/\theta}.$$

For the upper tail, l'Hôpital's rule and

$$\widetilde{\psi}'(t) = -\frac{\alpha}{\theta} t^{\alpha-1}(1+t^{\alpha})^{-(1+1/\theta)}$$

give

$$\begin{aligned}\widetilde{\lambda}_U &= 2 - \lim_{t\downarrow 0} \frac{1-\widetilde{\psi}(2t)}{1-\widetilde{\psi}(t)} \\ &= 2 - 2\lim_{t\downarrow 0} \frac{\widetilde{\psi}'(2t)}{\widetilde{\psi}'(t)} = 2 - 2^{\alpha}.\end{aligned}$$

□

Example 21.3 For $\lambda \in [0,1]$, define

$$C_{\lambda}(u_1,u_2) = \lambda M(u_1,u_2) + (1-\lambda)W(u_1,u_2). \tag{21.3}$$

Compute the following quantities.

1. Compute Spearman's rho $\rho_S(C_{\lambda})$.
2. Compute Kendall's tau $\tau(C_{\lambda})$.
3. Compute $\lambda_L(C_{\lambda})$ and $\lambda_U(C_{\lambda})$.

The mixture has mass on the two diagonals of the unit square, as illustrated in Figure 20.

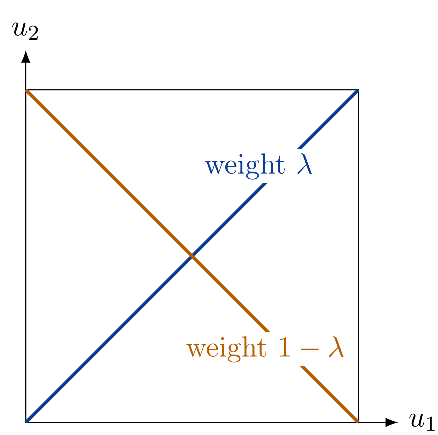


Figure 20

Solution. Let U be uniformly distributed on $[0,1]$. A random vector with copula C_λ has the mixture representation

$$(U_1, U_2) \stackrel{\mathrm{d}}{=} \begin{cases} (U, U), & \text{with probability } \lambda, \\ (U, 1-U), & \text{with probability } 1-\lambda. \end{cases} \tag{21.4}$$

For Part (1),

$$\mathbb{E}[U_1 U_2] = \lambda \mathbb{E}[U^2] + (1-\lambda)\mathbb{E}[U(1-U)] = \frac{\lambda}{3} + \frac{1-\lambda}{6}.$$

Proposition 12.5(1) therefore gives

$$\rho_S(C_\lambda) = 12\mathbb{E}[U_1 U_2] - 3 = 2\lambda - 1. \tag{21.5}$$

For Part (2), first use (21.4) to calculate

$$\mathbb{E}[M(U_1, U_2)] = \lambda \mathbb{E}[U] + (1-\lambda)\mathbb{E}[\min\{U, 1-U\}] = \frac{1+\lambda}{4},$$
$$\mathbb{E}[W(U_1, U_2)] = \lambda \mathbb{E}[(2U-1)_+] = \frac{\lambda}{4}.$$

It follows that

$$\mathbb{E}[C_\lambda(U_1, U_2)] = \lambda \frac{1+\lambda}{4} + (1-\lambda)\frac{\lambda}{4} = \frac{\lambda}{2}.$$

Proposition 12.5(2) now gives

$$\tau(C_\lambda) = 4\mathbb{E}[C_\lambda(U_1, U_2)] - 1 = 2\lambda - 1. \tag{21.6}$$

For Part (3), when $q < 1/2$,

$$C_\lambda(q, q) = \lambda q,$$

and hence

$$\lambda_L(C_\lambda) = \lim_{q \downarrow 0} \frac{C_\lambda(q, q)}{q} = \lambda.$$

When $q > 1/2$,

$$C_\lambda(q, q) = \lambda q + (1-\lambda)(2q - 1).$$

Consequently,

$$\lambda_U(C_\lambda) = \lim_{q \uparrow 1} \frac{1 - 2q + C_\lambda(q, q)}{1 - q} = \lambda.$$

This equality of the two tail coefficients also follows from radial symmetry of both M and W. $\square$

Appendix: Lecture and Tutorial Index