A method for constructing multivariate copulas

F. Durante
Department of Knowledge-Based
Mathematical Systems
Johannes Kepler University
A-4040 Linz, Austria
fabrizio.durante@jku.at

J.J. Quesada-Molina
Departamento de
Matemática Aplicada
Universidad de Granada
E-18071 Granada, Spain
jquesada@ugr.es

M. Úbeda-Flores
Departamento de Estadística
y Matemática Aplicada
Universidad de Almería
E-04120 Almería, Spain
mubeda@ual.es

Abstract

We provide a method for constructing a class of multivariate copulas depending on a univariate function. We study some properties of this class and present several examples. The same circle of ideas is used in a similar construction of quasi–copulas.

Keywords: Copula, Quasi–copula, Concordance, Kendall's tau.

1 Introduction

An n-dimensional *copula* (briefly n-copula) is a function $C\colon \mathbb{I}^n \to \mathbb{I}$ ($\mathbb{I} = [0,1]$) which satisfies:

(C1) for every $\mathbf{u} = (u_1, u_2, \ldots, u_n)$ in $\mathbb{I}^n$, $C(\mathbf{u}) = 0$ if at least one coordinate of $\mathbf{u}$ is 0, and $C(\mathbf{u}) = u_k$ whenever all coordinates of $\mathbf{u}$ are 1 except u_k;

(C2) for every $\mathbf{a} = (a_1, a_2, \ldots, a_n)$ and $\mathbf{b} = (b_1, b_2, \ldots, b_n)$ in $\mathbb{I}^n$ such that $a_k \le b_k$ for all $k = 1, 2, \ldots, n$, $V_C([\mathbf{a}, \mathbf{b}]) = \sum \mathrm{sgn}(\mathbf{c}) C(\mathbf{c}) \ge 0$, where $[\mathbf{a}, \mathbf{b}]$ denotes the *n-box* $[a_1, b_1] \times [a_2, b_2] \times \cdots \times [a_n, b_n]$, the sum is taken over all the *vertices* $\mathbf{c} = (c_1, c_2, \ldots, c_n)$ of $[\mathbf{a}, \mathbf{b}]$, $c_k \in \{a_k, b_k\}$ ($1 \le k \le n$), and sgn($\mathbf{c}$)= 1 if $c_k = a_k$ for an even number of indices k's, and sgn($\mathbf{c}$)= -1 if $c_k = a_k$ for an odd number of indices k's.

In view of *Sklar's Theorem* [15], the joint distribution function H of the random vector $(X_1, X_2, \ldots, X_n)$ with univariate marginals $F_1, F_2, \ldots, F_n$ can be expressed, for every $\mathbf{x} \in \mathbb{R}^n$, by

$$H(\mathbf{x}) = C(F_1(x_1), F_2(x_2), \ldots, F_n(x_n)), \qquad (1)$$

where C is an n-copula that is uniquely determined on $\mathrm{Range}\, F_1 \times \mathrm{Range}\, F_2 \times \cdots \times \mathrm{Range}\, F_n$.

Conversely, given n univariate distribution functions $F_1, F_2, \ldots, F_n$, and an n–copula C, the function H given by (1) is an n–dimensional distribution function.

In particular, Π_n and M_n, defined for all $\mathbf{u}$ in $\mathbb{I}^n$ by $\Pi_n(\mathbf{u}) = \prod_{i=1}^{n} u_i$ and $M_n(\mathbf{u}) = \min\{u_1, u_2, \ldots, u_n\}$, are the n-copula of independent and comonotone random variables, respectively.

Thus, copulas are useful tools in the construction of multivariate distributions with given marginals: it suffices to construct a multivariate copula and, hence, attach to it some univariate marginals.

In [8, 12], several methods for constructing copulas are given; however, most of them concern the bivariate case and no simple extension to the n–dimensional case ($n \ge 3$) is provided.

Recently [4], F. Durante introduced a new family of bivariate copulas depending on a univariate function. Specifically, under some assumptions on a function $f\colon \mathbb{I} \to \mathbb{I}$, the family of copulas given by

$$C_f(x, y) = \min\{x, y\} f(\max\{x, y\})$$

is studied (for more details about this family see [5]). In [6], it was proved that this family arises in a natural way when we impose that the horizontal and vertical sections of a bivariate copula are linear on some segments of the unit square.

In this note, we extend the above family of copulas to the n–dimensional case ($n \ge 3$) and we study its properties. More details can be found in [7].

2 The new family of n-copulas

Given a continuous function $f\colon \mathbb{I} \to \mathbb{I}$, we define the function $C_f : \mathbb{I}^n \to \mathbb{I}$ given by

$$C_f(u_1, u_2, \ldots, u_n) = u_{[1]} \prod_{i=2}^{n} f(u_{[i]}) \qquad (2)$$

where $u_{[1]}, \dots, u_{[n]}$ denote the components of $(u_1, u_2, \dots, u_n) \in \mathbb{I}^n$ rearranged in increasing order. We aim at studying the conditions under which C_f is a copula.

Theorem 1. *Let $f\colon \mathbb{I} \longrightarrow \mathbb{I}$ be a continuous function and let C_f be the function defined by (2). Then C_f is an n–copula if, and only if,*

(a) *$f(1) = 1$;*

(b) *f is increasing;*

(c) *the function $t \to f(t)/t$ is decreasing on $(0, 1]$.*

A function f satisfying the assumptions of Theorem 1 is called a *generator*. Every generator f is the restriction to $\mathbb{I}$ of a univariate distribution function. In particular, if $(U_1, U_2, \dots, U_n)$ is a vector of n random variables uniformly distributed on $\mathbb{I}$ with n-copula C_f, then

$$P(\max\{U_1, U_2, \dots, U_n\} \le t | U_1 \le t) = f^{n-1}(t).$$

Note that if $f_1(t) = t$ for all $t \in \mathbb{I}$, then $C_{f_1} = \Pi_n$; and, if $f_2(t) = 1$ for all $t \in \mathbb{I}$, then $C_{f_2} = M_n$. In general, if f is a generator, then $t \le f(t) \le 1$. Moreover, $f(t)/t$ is decreasing on $(0, 1]$ if, and only if, f is star–shaped, viz. $f(\alpha t) \ge \alpha f(t)$ for all α, t in $\mathbb{I}$; in particular, if f is concave, then $f(t)/t$ is decreasing on $(0, 1]$.

In the sequel, we will denote by Φ the class of all generators. When building n-copulas of type (2), the class Φ plays a major rôle. Some properties of this class are presented in the following result.

Proposition 1. *Let f and g be two continuous functions from $\mathbb{I}$ onto $\mathbb{I}$. The following statements hold:*

(a) *if f and g are in Φ, then $\alpha f + (1 - \alpha) g$ is in Φ for every α in $\mathbb{I}$;*

(b) *if f and g are in Φ, then the functions $\min\{f, g\}$ and $\max\{f, g\}$ are in Φ;*

(c) *if f and g are in Φ, then the composition $f \circ g$ is in Φ.*

Example 1. For any α in $\mathbb{I}$, consider the function $f\colon \mathbb{I} \to \mathbb{I}$ given by $f_\alpha(t) = \alpha t + \overline{\alpha}$, with $\overline{\alpha} := 1 - \alpha$. Then, f_α is in Φ, and the n-copula $C_{f_\alpha} = C_\alpha$ defined by (2) is given by

$$C_\alpha(\mathbf{u}) = u_{[1]} \prod_{i=2}^{n} (\alpha u_{[i]} + \overline{\alpha}).$$

In particular, in the bivariate case, we obtain the well–known Fréchet family of copulas $C_\alpha(u_1, u_2) = \alpha u_1 u_2 + (1 - \alpha) \min\{u_1, u_2\}$ (see [8, Family B11] and [12]).

Example 2. For any $\alpha \ge 1$, consider the function $f\colon \mathbb{I} \to \mathbb{I}$ given by $f_\alpha(t) := \min\{\alpha t, 1\}$. Then, f_α is in Φ, and the n-copula $C_{f_\alpha} = C_\alpha$ defined by (2) is given by

$$C_\alpha(\mathbf{u}) = u_{[1]} \prod_{i=2}^{n} \min\{\alpha u_{[i]}, 1\}.$$

In particular, in the bivariate case, C_α is the *ordinal sum* of the 2-copulas $\{\Pi_2, M_2\}$ with respect to the partition $\{[0, 1/\alpha], [1/\alpha, 1]\}$ (see [12] for more details).

Example 3. For any α in $\mathbb{I}$, consider the function $f\colon \mathbb{I} \to \mathbb{I}$ given by $f_\alpha(t) = t^\alpha$. Then, f_α is in Φ, and the n-copula $C_{f_\alpha} = C_\alpha$ defined by (2) is given by

$$C_\alpha(\mathbf{u}) = u_{[1]}^{1-\alpha} \prod_{i=1}^{n} u_i^\alpha,$$

i.e., $C_\alpha = (M_n)^{1-\alpha} (\Pi_n)^\alpha$. Note that $C_0 = M_n$ and $C_1 = \Pi_n$. Moreover, C_α can be considered as a generalization of the Cuadras-Augé family of 2-copulas [2]. In this case, every copula C_α is a *multivariate extreme copula*, viz. $C_\alpha(u_1^t, u_2^t, \dots, u_n^t) = (C_\alpha(u_1, u_2, \dots, u_n))^t$ for every $t > 0$, which is a useful property in multivariate extreme value theory, as showed in [8].

Notice that n–copula C_f is *symmetric*, viz. the value of C_f does not change by permuting its arguments.

The mixed derivative of order n of an n–copula C, $\frac{\partial^n C}{\partial u_1 \dots \partial u_n}(\mathbf{u})$, exists almost everywhere on $\mathbb{I}^n$. In particular, an n–copula C is said to be *absolutely continuous* if

$$C(\mathbf{u}) = \int_0^{u_1} \cdots \int_0^{u_n} \frac{\partial^n C(\mathbf{t})}{\partial t_1 \dots \partial t_n} \, d\mathbf{t},$$

otherwise, C has a *singular component*. Every n–copula C_f of type (2), except Π_n, has a singular component. In fact, as illustrated in [8, pages 14–15], it suffices to note that, for every $i = 1, 2, \dots, n$, the mapping $t \mapsto \frac{\partial C}{\partial u_i}(u_1, \dots, u_{i-1}, t, u_{i+1} \dots, u_n)$ has a jump discontinuity. For instance, in the bivariate case, the first derivative of C_f is given by

$$\frac{\partial C_f}{\partial u}(u, v) = \begin{cases} f(v), & \text{if } u < v, \\ v\, f'(u), & \text{otherwise.} \end{cases}$$

For a fixed v_0, the mapping $t \mapsto \frac{\partial C_f}{\partial u}(t, v_0)$ has a jump discontinuity in v_0, and, thus, C_f has a singular component along the main diagonal of the unit square. By using [8, Theorem 1.1], the mass of this singular component is given by

$$m = \int_0^1 (f(t) - t f'(t)) \, dt = 2 \int_0^1 f(t) \, dt \, - 1.$$

This m has a graphical interpretation when f admits an inverse; in fact, m is the area of the region of the unit square between the graph of f and the graph of f^{-1}.

3 Statistical properties

Now, we give a statistical interpretation of the new family. Let $W_1, W_2, \dots, W_n$, Z be $n+1$ independent random variables such that W_i has distribution function f satisfying parts (a), (b) and (c) in Theorem 1, for all $i = 1, 2, \dots, n$, and Z has distribution function $g(t) = t/f(t)$. Note that $g(1) = 1$ and g is increasing, since $f(t)/t$ is decreasing. Consider the random variables $U_i = \max\{W_i, Z\}$, for all $i = 1, 2, \dots, n$. Then, for every $(u_1, u_2, \dots, u_n)$, the distribution function of the random vector $(U_1, U_2, \dots, U_n)$ is given by

$$P(U_1 \leq u_1, \dots, U_n \leq u_n) = u_{[1]} \prod_{i=2}^{n} f(u_{[i]}),$$

and, hence, it is a copula of type (2).

From a statistical point of view, the study of concordance in a family of multivariate distributions has also a great interest: this is the topic of the following result. We recall that, given two n-copulas C_1 and C_2, C_1 is said to be *more concordant than* C_2 (written $C_1 \succ C_2$) if both $C_1 \geq C_2$ and $\overline{C}_1 \geq \overline{C}_2$ hold, where, if $\mathbf{U}$ is a random vector with joint d.f. given by the n-copula C, then $\overline{C}$ is the *survival function* associated with C defined by $\overline{C}(\mathbf{u}) = P[\mathbf{U} > \mathbf{u}]$. For more details see [8, 12].

Theorem 2. *Let f and g be two generators and let C_f and C_g be two n–copulas of type* (2) *defined by $C_f(\mathbf{u}) = u_{[1]} \prod_{i=2}^{n} f(u_{[i]})$ and $C_g(\mathbf{u}) = u_{[1]} \prod_{i=2}^{n} g(u_{[i]})$, respectively, for every $\mathbf{u} = (u_1, u_2, \dots, u_n) \in \mathbb{I}^n$. Then, we have that $C_f \prec C_g$ if, and only if, both the following conditions are satisfied:*

(i) $\prod_{i=2}^{n} f(u_{[i]}) \leq \prod_{i=2}^{n} g(u_{[i]})$ *for all* $\mathbf{u}$ *in* $\mathbb{I}^n$, *and*

(ii) *for all* $\mathbf{u}$ *in* $\mathbb{I}^n$

$$\sum_{i=2}^{n} (f(u_{[i]}) - g(u_{[i]})) + (1 - u_{[1]}) \cdot \left(\prod_{i=2}^{n} (1 - f(u_{[i]})) - \prod_{i=2}^{n} (1 - g(u_{[i]})) \right) \leq 0. \quad (3)$$

In particular, for $n = 2$, we have that $C_f \prec C_g$ if, and only if, $f \leq g$.

Another way to summarize the information about the concordance for copulas is represented by the so–called *multivariate measures of concordance* ([16]), which are non-parametric measures of multivariate association for a continuous random vector with associated n-copula C. The most common measure is Kendall's tau (see [10, 11]), which is given by

$$\tau_n(C) = \frac{1}{2^{n-1} - 1} \left(2^n \int_{\mathbb{I}^n} C(\mathbf{u}) \, \mathrm{d}C(\mathbf{u}) - 1 \right).$$

For a copula of type (2), it has the following expression.

Theorem 3. *Let C_f be the n–copula given by* (2) *via Theorem* 1*. Then, Kendall's tau associated with C_f is given by:*

$$\tau_n(C_f) = \frac{1}{2^{n-1} - 1} \left(2^{n-1} - 2^n \sum_{k=2}^{n} \frac{k}{2^{k-1}} \int_{\mathbb{I}} t(1 - f^2(t))^{k-1} \, \mathrm{d}t - 1 \right).$$

Example 4. Consider the copula C_α of Example 3. The value of Kendall's tau for C_α is given by:

$$\tau_n(C) = \frac{1}{2^{n-1} - 1} \left(2^{n-1} - \frac{2^n}{\alpha} \sum_{k=2}^{n} \frac{k}{2^k} \operatorname{Beta}\left(\frac{1}{\alpha}, k \right) - 1 \right).$$

4 A new family of n-quasi–copulas

The notion of *quasi–copula* was introduced by Alsina, Nelsen and Schweizer [1] in order to show that a certain class of operations on univariate distribution functions is not derivable from corresponding operations on random variables defined on the same probability space (see also [13] for the multivariate case). Cuculescu and Theodorescu [3] characterized an n-dimensional quasi–copula (or n-quasi–copula) as a function Q from $\mathbb{I}^n$ onto $\mathbb{I}$, which satisfies (C1), and, in place of (C2), both the weaker conditions:

(Q1) Q is *non-decreasing in each variable*;

(Q2) Q satisfies the *Lipschitz condition*,

$$|Q(\mathbf{u}) - Q(\mathbf{v})| \leq \sum_{i=1}^{n} |u_i - v_i|$$

for all $\mathbf{u}$, $\mathbf{v}$ in $\mathbb{I}^n$.

While every n-copula is an n-quasi–copula, there exist *proper* n-quasi–copulas, i.e., n-quasi–copulas which are not n-copulas. For example,

$$W_n(\mathbf{u}) = \max\left\{ 0, \sum_{i=1}^{n} u_i - n + 1 \right\}$$

is an n–copula if, and only if, $n = 2$, and a proper n-quasi–copula for all $n \geq 3$. Recently, n-quasi–copulas have been used to express the pointwise best-possible bounds on nonempty sets of distribution functions, n-copulas or n-quasi–copulas (see [14]). Here we present the characterization of quasi–copulas of type (2).

Theorem 4. *Let $f : \mathbb{I} \to \mathbb{I}$ be a continuous function, and let C_f be the function defined by (2). Then, C_f is an n–quasi–copula if, and only if, the following statements are satisfied:*

(i) $f(1) = 1$,

(ii) f *is increasing,*

(iii) $u_1(f(u_2) - f(u_1)) \leq u_2 - u_1$ *for every* $u_1, u_2 \in \mathbb{I}$, *with* $u_1 < u_2$.

Proof. If C_f is an n-quasi–copula, then it is easy to check that (i) and (ii) hold.

To prove (iii), consider $u_1, v_1 \in \mathbb{I}$ with $u_1 < v_1$. From (Q2) we have that

$$C_f(v_1, u_1, 1, \dots, 1) - C_f(u_1, u_1, 1, \dots, 1) \leq v_1 - u_1,$$

that is $u_1(f(v_1) - f(u_1)) \leq v_1 - u_1$, and we obtain (iii).

Conversely, given a continuous $f : \mathbb{I} \to \mathbb{I}$ satisfying (i), (ii) and (iii), let C_f be the function defined by (2). Then it is easily proved that C_f satisfies (C1) and (Q1). In order to prove that C_f satisfies (Q2), let $u_1, v_1, \dots, v_n$ be $(n+1)$ points in $\mathbb{I}$ such that $u_1 < v_1$. Notice that, because C_f is symmetric, it is enough to show that

$$\lambda = C_f(v_1, v_2, \dots, v_n) - C_f(u_1, v_2, \dots, v_n) \leq v_1 - u_1. \quad (4)$$

Three cases will be considered.

1. If $v_{[1]} = v_1$, we obtain that

$$\lambda = (v_1 - u_1) \prod_{i=2}^{n} f(v_i),$$

and therefore (4) holds.

2. If $v_{[1]} = v_j \leq u_1$ for some $j \in \{2, 3, \dots, n\}$, then we have that

$$\lambda = v_j(f(v_1) - f(u_1)) \prod_{\substack{i=2 \\ i \neq j}}^{n} f(v_i),$$

and (4) follows from (iii) and the fact that $f(t) \leq 1$ for all $t \in \mathbb{I}$.

3. Finally, if $u_1 < v_{[1]} = v_j < v_1$ for some j belongs to $\{2, 3, \dots, n\}$, then

$$\lambda = [C_f(v_j, v_2, \dots, v_n) - C_f(u_1, v_2, \dots, v_n)] + [C_f(v_1, v_2, \dots, v_n) - C_f(v_j, v_2, \dots, v_n)].$$

Inequality (4) is, hence, a consequence of the preceding two cases.

Hence, the proof is completed. □

Corollary 1. *Let $f : \mathbb{I} \longrightarrow \mathbb{I}$ be a differentiable function, and let C_f be the function defined by (2). Then, C_f is an n-quasi–copula if, and only if, the following statements are satisfied:*

(i) $f(1) = 1$;

(ii) f *is increasing;*

(iii) $uf'(u) \leq 1$ *for every* $u \in \mathbb{I}$.

We now provide an example of a proper n-quasi–copula of type (2).

Example 5. Consider the function $f(u) = u + u^2 - u^3$ for every $u \in \mathbb{I}$, and let C_f be the function defined by (2). Then, it is easy to check that $f(1) = 1$, f is increasing on $[0, 1]$, but C_f is not an n-copula since the function $f(u)/u$ is increasing on $[0, 1/2]$. However, it is easy to see that $uf'(u) = u(1 + 2u - 3u^2)$ satisfies part (iii) in Corollary 1, and thus, C_f is a proper n-quasi–copula.

5 Concluding remarks

We have introduced a new family of multivariate copulas depending on a univariate function. Now, we will show how to construct many other copulas starting from our method and a result from [9].

Consider a continuous and increasing bijection $\phi : \mathbb{I} \to \mathbb{I}$, and suppose that ϕ^{-1} is absolutely monotonic of order n on $\mathbb{I}$, viz. ϕ^{-1} admits derivatives up to order n on $\mathbb{I}$ and, for $i = 1, 2, \dots, n$,

$$\frac{\mathrm{d}^i(\phi^{-1})(t)}{\mathrm{d}t^i} \geq 0.$$

From [9, Theorem 4.7], for every n–copula C, we have that the mapping $C_\phi : \mathbb{I}^n \to \mathbb{I}$ defined by

$$C_\phi(u_1, \dots, u_n) = \phi^{-1}(C(\phi(u_1), \dots, \phi(u_n))) \quad (5)$$

is also an n–copula.

In particular, if $C = C_f$ is a copula of type (2) generated by f, then the function $C_{f,\phi}$ defined by

$$C_{f,\phi}(u_1, u_2, \dots, u_n) = \phi^{-1}\left[\phi(u_{[1]}) \prod_{i=2}^{n} f(\phi(u_{[i]}))\right]$$

is also an n–copula.

For $g_1 = -\ln \phi$ and $g_2 = -\ln(f \circ \phi)$, equation (5) can be written into the form

$$C_{g_1,g_2}(u_1, u_2, \dots, u_n) = g_1^{-1}\left[g_1(u_{[1]}) + \sum_{i=2}^{n} g_2(u_{[i]})\right].$$

The reader will recognize that the last expression is a direct generalization of the Archimedean family of multivariate copulas [12].

Acknowledgement

The second and third authors acknowledge the support by the Ministerio de Educación y Ciencia (Spain) and FEDER, under research project MTM2006-12218.

References

[1] C. Alsina, R.B. Nelsen, B. Schweizer, On the characterization of a class of binary operations on distribution functions, Statist. Probab. Lett. 17 (1993) 85–89.

[2] C.M. Cuadras, J. Augé, A continuous general multivariate distribution and its properties, Comm. Statist. - Theory Methods 10 (1981) 339–353.

[3] I. Cuculescu, R. Theodorescu, Copulas: diagonals and tracks, Rev. Roumaine Math. Pures Appl. 46 (2001) 731–742.

[4] F. Durante, A new family of symmetric bivariate copulas, C.R. Acad. Sci. Paris, Sér. I, 344 (2007) 195–198.

[5] F. Durante, A new class of symmetric bivariate copulas, J. Nonparametr. Stat., in press.

[6] F. Durante, R. Mesiar, A. Kolesárová, C. Sempi, Semilinear copulas, submitted.

[7] F. Durante, J.J. Quesada-Molina, M. Úbeda-Flores, On a family of multivariate copulas for aggregation processes, submitted.

[8] H. Joe, Multivariate Models and Dependence Concepts, Chapman & Hall, New York, 1997.

[9] P.M. Morillas, A method to obtain new copulas from a given one, Metrika 61 (2005) 169–184.

[10] R.B. Nelsen, Nonparametric measures of multivariate association, in: L. Rüschendorf, B. Schweizer, M.D. Taylor (Eds.) Distribution Functions with Fixed Marginals and Related Topics, Institute of Mathematical Statistics (Lecture Notes – Monograph Series Volume 28), Hayward CA, 1996, pp. 223–232.

[11] R.B. Nelsen, Concordance and copulas: a survey, in: C. Cuadras, J. Fortiana, J.A. Rodríguez-Lallena (Eds.) Distributions with Given Marginals and Statistical Modelling, Kluwer Academic Publishers, Dordrecht, 2002, pp. 169–178.

[12] R.B. Nelsen, An Introduction to Copulas, Springer, New York, 2006.

[13] R.B. Nelsen, J.J. Quesada-Molina, B. Schweizer, C. Sempi, Derivability of some operations on distribution functions, in: L. Rüschendorf, B. Schweizer and M.D. Taylor (Eds.), Distributions with Fixed Marginals and Related Topics, IMS Lecture Notes-Monograph Series, Hayward, CA, 1996, pp. 233–243.

[14] J.A. Rodríguez-Lallena, M. Úbeda-Flores, Best-possible bounds on sets of multivariate distribution functions, Comm. Statist. - Theory Methods 33 (2004) 805–820.

[15] A. Sklar, Fonctions de répartition à n dimensions et leurs marges, Publ. Inst. Statist. Univ. Paris 8 (1959) 229–231.

[16] M.D. Taylor, Multivariate measures of concordance, Ann. Inst. Statist. Math., in press.