Monografías del Semin. Matem. García de Galdeano. **27:** 499–506, (2003).

What are copulas?

José Juan Quesada-Molina

Departamento de Matemática Aplicada, Universidad de Granada

José Antonio Rodríguez-Lallena

Manuel Úbeda-Flores

Departamento de Estadística y Matemática Aplicada, Universidad de Almería

Abstract

The notion of copula was introduced by A. Sklar in 1959, when answering a question raised by M. Fréchet about the relationship between a multidimensional probability function and its lower dimensional margins. At the beginning, copulas were mainly used in the development of the theory of probabilistic metric spaces. Later, they were of interest to define nonparametric measures of dependence between random variables, and since then, they began to play an important role in probability and mathematical statistics. In this paper, a general overview of the theory of copulas will be presented. Some of the main results of this theory, various examples, and some open problems will be described.

MSC: Primary 60E05; Secondary 62H05, 62H20.

Keywords: Copulas; dependence concepts; measures of association, probabilistic metric spaces.

1 Introduction

During a long time statisticians have been interested on the relationship between a multivariate distribution function and its lower dimensional margins (univariate or of higher dimensions). M. Fréchet (see [11]), and G. Dall'Aglio (see [6]) did some interesting works about this matter in the fifties, studying the bivariate and trivariate distribution functions with given univariate margins. The answer to this problem for the univariate margins case was given by A. Sklar in 1959 (see [31]) creating a new class of functions which he called *copulas*. These new functions are restrictions to $[0,1]^2$ of bivariate distribution functions whose margins are uniform in $[0,1]$. In short, Sklar showed that if H is a bivariate distribution function with margins $F(x)$ and $G(y)$, then there exists a copula C such that $H(x,y) = C(F(x), G(y))$.

Between 1959 and 1976 most of the results about copulas were obtained in the course of the development of the probabilistic metric spaces, mainly in the study of binary operations in the space of the probability distribution functions. In 1942, Karl Menger (see [20]) proposed a probabilistic generalization of the theory of metric spaces, by replacing the number $d(p,q)$ by a distribution function F_{pq}, whose value $F_{pq}(x)$ for any real x is the probability that the distance between p and q is less than x. The first difficulty in the construction of probabilistic metric spaces comes when one tries to find a "probabilistic" analog of the triangle inequality. Menger proposed $F_{pr}(x+y) \geq T(F_{pq}(x), F_{qr}(y))$, where T is a *triangle norm* or *t-norm*. Some t-norms are copulas, and conversely, some copulas are t-norms. For a history of the development of the theory of probabilistic metric spaces, see [28] and [29].

Subsequently, it was discovered that copulas could be useful to define nonparametric measures of dependence between random variables. Since then, the concept of copula has been rediscovered in several times, playing an important role in Probability and Statistics, particularly in problems related to dependence, given marginals and functions of random variables that are invariants under monotone transformations. A historical review about the evolution of this matter can be found in [7] and [28]. The recent book by R.B. Nelsen (see [21]) is an important monograph about copulas. As for the relationship with problems of given marginals, it can be seen [2], [5], [8] and [27].

2 Copulas

We begin with the definition of copula for the bivariate case.

Definition 2.1. A *copula* is a function $C : [0,1]^2 \longrightarrow [0,1]$ which satisfies:

(a) For every u, v in $[0,1]$, $C(u,0) = 0 = C(0,v)$, and $C(u,1) = u$ and $C(1,v) = v$;

(b) for every u_1, u_2, v_1, v_2 in $[0,1]$ such that $u_1 \leq u_2$ and $v_1 \leq v_2$, $C(u_2,v_2) - C(u_2,v_1) - C(u_1,v_2) + C(u_1,v_1) \geq 0$.

The importance of copulas in Statistics is described in Sklar's Theorem:

Theorem 2.1. *Let X and Y be random variables with joint distribution function H and marginal distribution functions F and G, respectively. Then there exists a copula C such that*

$$H(x,y) = C(F(x), G(y)) \tag{1}$$

for all x, y in $\mathbb{R}$. If F and G are continuous, then C is unique. Otherwise, the copula C is uniquely determined on $Ran(F) \times Ran(G)$. Conversely, if C is a copula and F and

G are distribution functions, then the function H defined by (1) is a joint distribution function with margins F and G.

Thus copulas link joint distribution functions to their one-dimensional margins. A proof of this theorem can be found in [29].

A first example of copulas is the product copula $\Pi(u,v) = uv$, which characterizes independent random variables when the distribution functions are continuous.

As a consequence of Sklar's Theorem, we encounter the Fréchet-Hoeffding bounds for copulas, i.e., for any copula C and for all u, v in $[0,1]$,

$$W(u,v) = \max(u+v-1, 0) \leq C(u,v) \leq \min(u,v) = M(u,v),$$

where W and M are also copulas.

Much of the usefulness of copulas in the study of nonparametric statistics derives from the facts expressed in the following result.

Theorem 2.2. *Let X and Y be continuous random variables with copula C_{XY}. Let f and g be strictly monotone functions on Ran(X) and Ran(Y), respectively.*

(a) *If f and g are strictly increasing, then $C_{f(X),g(Y)}(u,v) = C_{XY}(u,v)$.*

(b) *If f is strictly increasing and g is strictly decreasing, then $C_{f(X),g(Y)}(u,v) = u - C_{XY}(u, 1-v)$.*

(c) *If f is strictly decreasing and g is strictly increasing, then $C_{f(X),g(Y)}(u,v) = v - C_{XY}(1-u, v)$.*

(d) *If f and g are strictly decreasing, then $C_{f(X),g(Y)}(u,v) = u+v-1+C_{XY}(1-u, 1-v)$.*

3 Examples of families of copulas

In this section we show some of the most known family of copulas.

Example 3.1. Fréchet's family (1958) (see [12]). The following two-parameter family of copulas is a convex linear combination of the copulas Π, W and M, $C_{\alpha,\beta}(u,v) = \alpha M(u,v) + (1-\alpha-\beta)\Pi(u,v) + \beta W(u,v)$, where $\alpha, \beta \in [0,1]$ with $\alpha + \beta \leq 1$.

Example 3.2. Farlie-Gumbel-Morgenstern's family (1960) (see [10]). If $\theta \in [-1,1]$, then the function $C_\theta(u,v) = uv + \theta uv(1-u)(1-v)$ is a one-parameter family of copulas. A generalization of this family can be found in [26].

Example 3.3. Marshall-Olkin's family (1967) (see [19]). If $\alpha, \beta \in [0,1]$, then $C_{\alpha,\beta}(u,v) = \min(u^{1-\alpha}v, uv^{1-\beta})$ is a two-parameter family of copulas.

Example 3.4. Archimedean copulas. Copulas of the form $C(u,v) = \varphi^{[-1]}(\varphi(u) + \varphi(v))$ are called *Archimedean* copulas, where $\varphi^{[-1]}$ is the pseudo-inverse of a continuous and strictly decreasing function from $[0,1]$ to $[0,\infty]$ with $\varphi(1) = 0$. For a detailed study of these copulas, see [13].

4 Copulas and association

In this section we will look at different ways in which copulas can be used in the study of dependence between random variables. For a historical review of measures of association and concepts of dependence, see [17] and [18]. For some recent results, see [4], [21], [22], and [30].

4.1 Measures of association

4.1.1 Kendall's τ

Kendall's tau measure of a pair (X, Y), distributed according to H, can be defined as the difference between the probabilities of concordance and discordance for two independent pairs (X_1, Y_1) and (X_2, Y_2) each with distribution H; that is

$$\tau_{XY} = \Pr\{(X_1 - X_2)(Y_1 - Y_2) > 0\} - \Pr\{(X_1 - X_2)(Y_1 - Y_2) < 0\}.$$

These probabilities can be evaluated by integrating over the distribution of (X_2, Y_2). So that, in terms of copulas, Kendall's τ becomes to

$$\tau_C = 4 \int_0^1 \int_0^1 C(u, v) dC(u, v) - 1,$$

where C is the copula associated to (X, Y).

4.1.2 Spearman's ρ

Let (X_1, Y_1), (X_2, Y_2) and (X_3, Y_3) be three independent random vectors with a common joint distribution function H. Consider the vectors (X_1, Y_1) and (X_2, Y_3), then the Spearman's ρ coefficient associated to a pair (X, Y), distributed according to H, is defined as

$$\rho_{XY} = 3(\Pr\{(X_1 - X_2)(Y_1 - Y_3) > 0\} - \Pr\{(X_1 - X_2)(Y_1 - Y_3) < 0\}).$$

In terms of the copula C associated to the pair (X, Y) becomes to

$$\rho_C = 12 \int_0^1 \int_0^1 (C(u, v) - uv) dudv. \tag{2}$$

4.1.3 Schweizer and Wolff's σ

If we replace the function $C(u, v) - uv$ in (2) by its absolute value, then we obtain Schweizer and Wolff's σ, given by

$$\sigma_C = 12 \int_0^1 \int_0^1 |C(u, v) - uv| dudv.$$

4.2 Dependence concepts

Let X and Y be continuous random variables with joint distribution function H and marginals F and G, respectively. We say that X and Y are *positive quadrant dependent* if $H(x, y) - F(x)G(y) \geq 0$ for all $x, y \in \mathbb{R}$. In terms of the copula C_{XY} associated to the pair (X, Y), it means that $C_{XY}(u, v) \geq uv$ for all $u, v \in [0, 1]$.

See [15] and [16] for further discussions of this concept of dependence and many others.

Lastly, we note that there exist some relationships among some measures of association and certain dependence concepts. For a complete review, see [21].

5 Other problems related to copulas

5.1 Operations on distribution functions and quasi-copulas

A binary operation χ on the set of distribution functions is *derivable* from a function on random variables if there exists a Borel-measurable two-place function Z satisfying the following condition: For every pair of distribution functions F and G, there exist two random variables X and Y such that F and G are, respectively, the distribution functions of X and Y, and $\chi(F, G)$ is the distribution function of the random variable $Z(X, Y)$.

The notion of *quasi-copula* was introduced in [1] to characterize operations on distribution functions that can or cannot be derived from operations on random variables (see also [25]). Genest *et al.* (see [14]) have characterized the quasi-copula concept in simpler operational terms, as the following result asserts:

Theorem 5.1.1. *A function $Q : [0, 1]^2 \longrightarrow [0, 1]$ is a quasi-copula if and only if it satisfies:*

(i) $Q(0, x) = Q(x, 0) = 0$ *and* $Q(x, 1) = Q(1, x) = x$ *for all* x *in* $[0, 1]$*;*

(ii) $Q(x, y)$ *is nondecreasing in each of its arguments; and*

(iii) *the Lipschitz condition* $|Q(x_1, y_1) - Q(x_2, y_2)| \leq |x_1 - x_2| + |y_1 - y_2|$ *for all* x_1*,* x_2*,* y_1 *and* y_2 *in* $[0, 1]$.

Recently (see [24]), it has been proved a new simple characterization of quasi-copulas and some properties of these functions, all of them concerning the mass distribution of a quasi-copula. It has been showed that the features of this mass distribution can be quite different from that of a copula.

5.2 Markov processes

We begin this subsection with a "product" operation for copulas studied in [9].

Definition 5.2.1. Let C_1 and C_2 be copulas. The *product* of C_1 and C_2 is the function $C_1 * C_2$ from $[0,1]^2$ to $[0,1]$ given by

$$(C_1 * C_2)(u,v) = \int_0^1 \frac{\partial C_1}{\partial v}(u,t)\frac{\partial C_2}{\partial u}(t,v)dt.$$

As a first result, the authors (see [9]) showed that $C_1 * C_2$ is a copula, and their main result is the following theorem:

Theorem 5.2.1. *Let $\{X_t | t \in T\}$ be a stochastic process, and for each s, t in T, let C_{st} denote the copula of the random variables X_s and X_t. Then the following conditions are equivalent:*

(a) *The conditional distribution functions $P(x, s; y, t)$ satisfy the Chapman-Kolmogorov equations for all $s < u < t$ in T and almost all x, y in $\mathbb{R}$;*

(b) *For all $s < u < t$ in T, $C_{st} = C_{su} * C_{ut}$.*

This theorem yields a new technique for constructing Markov processes.

6 New problems

There are new problems in the study of the theory of copulas under three different points of view. These are:

- Stochastic orderings: See [3] and [23] for more details.
- Nonparametric Statistics: The use of copulas to define nonparametric hypothesis testing.
- Probability Theory: Developments in the theory of quasi-copulas.
- Numerical Analysis: Methods of approximation and interpolation in a given family of copulas from data provided by a bivariate random sample.

References

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

[2] Beněs, V. and Štěpán J., editors (1997), *Distributions with Given Marginals and Moment Problems* (Kluwer Academic Publishers, Dordrecht).

[3] Capéràa P. and Genest, C. (1990), "Concepts de dépendance et ordres stochastiques pour des lois bidimensionnelles," *Canad. J. Statist.* **18**, 315-326.

[4] Capéràa P. and Genest, C. (1993), "Spearman's ρ is larger than Kendall's τ for positively dependent random variables," *J. Nonparametr. Statist.* **2**, 183-194.

[5] Cuadras, C.; Fortiana, J. and Rodríguez Lallena, J. A., editors (2002), *Distributions with Given Marginals and Statistical Modelling* (Kluwer Academic Publishers, Dordrecht).

[6] Dall'Aglio, G. (1959), "Sulla compatibilità delle funzioni di ripartizione doppia," *Rend. Mat.* **18**, 385-413.

[7] Dall'Aglio, G. (1991), "Fréchet classes: the beginnings" in *Advances in Probability Distributions with Given Marginals* (G. Dall'Aglio, S. Kotz and G. Salinetti, eds.), pp. 1-12, Kluwer Academic Publishers, Dordrecht.

[8] Dall'Aglio, G.; Kotz, S. and Salinetti, G., editors (1991), *Advances in Probability Distributions with Given Marginals* (Kluwer Academic Publishers, Dordrecht).

[9] Darsow, W. F.; Nguyen, B. and Olsen, E. T. (1992), "Copulas and Markov Processes," *Illinois J. Mathematics* **36**, 600-642.

[10] Farlie, D. J. G. (1960), "The perfomance of some correlation coefficients for a general bivariate distribution," *Biometrika* **47**, 307-323.

[11] Fréchet, M. (1951), "Sur les tableaux de corrélation dont les marges sont données," *Ann. Univ. Lyon Sect. A* **9**, 53-77.

[12] Fréchet, M. (1958), "Remarque au sujet de la note précédente," *R. Acad. Sci. Paris* **246**, 2719-2720.

[13] Genest, C. and MacKay, J. (1986), "Copules archimédiennes et familles de lois bidimensionnelles dont les marges sont données," *Canad. J. Statist.* **14**, 145-159.

[14] Genest, C.; Quesada Molina, J. J.; Rodríguez Lallena, J. A. and Sempi, C. (1999), "A characterization of quasi-copulas," *J. Multivariate Anal.* **69**, 193-205.

[15] Hutchinson, T. P. and Lai, C. D. (1990), Continuous Bivariate Distributions, Emphasising Applications (Rumsby Scientific Publishing, Adelaide).

[16] Joe, H. (1997), *Multivariate Models and Dependence Concepts*, Chapman & Hall, London.

[17] Kruskal, W. H. (1958), "Ordinal measures of association," *J. Amer. Statist. Assoc.* **53**, 814-861.

[18] Lehmann, E. L. (1966), "Some concepts of dependence," *Ann. Math. Statist.* **37**, 1137-1153.

[19] Marshall, A. W. and Olkin, I. (1967), "A generalized bivariate exponential distribution," *J. Appl. Probability* **4**, 291-302.

[20] Menger, K. (1942), "Statistical metrics," *Proc. Nat. Acad. Sci. U.S.A.* **28**, 535-537.

[21] Nelsen, R. B. (1999), An Introduction to Copulas, Springer, New York.

[22] Nelsen, R. B.; Quesada Molina, J. J.; Rodríguez Lallena, J. A. y Úbeda Flores, M. (2001), "Bounds on bivariate distribution functions with given margins and measures of association", *Comm. Statist. –Theory and Methods* vol. **30**, number 6, 1155-1162.

[23] Nelsen, R. B.; Quesada Molina, J. J.; Rodríguez Lallena, J. A. y Úbeda Flores, M. (2001), "Distribution functions of copulas: a class of bivariate probability integral transforms", *Statist. Probab. Lett.* **54**, 277-282.

[24] Nelsen, R. B.; Quesada Molina, J. J.; Rodríguez Lallena, J. A. and Úbeda Flores, M. (2001), "Some new properties of quasi-copulas," in *Distributions with Given Marginals and Statistical Modelling* (C. Cuadras, J. Fortiana and J.A. Rodríguez, eds.), Kluwer Academic Publishers, Dordrecht.

[25] Nelsen, R. B.; Quesada Molina, J. J.; Schweizer, B. y Sempi, C. (1996), "Derivability of some operations on distribution functions," in *Distributions with Fixed Marginals and Related Topics* (L. Rüschendorf, B. Schweizer, M. D. Taylor, eds.), IMS Lecture Notes-Monograph Series Number 28, pp. 233-243, Hayward, California.

[26] Quesada Molina, J. J. and Rodríguez Lallena, J. A. (1995), "Bivariate copulas with quadratic sections," *J. Nonparametr. Statist.* **5**, 323-337.

[27] Rüschendorf, L.; Schweizer, B. and Taylor M. D., editors (1996), *Distributions with Fixed Marginals and Related Topics*, (Institute of Mathematical Statistics, Hayward, California).

[28] Schweizer, B. (1991), "Thirty years of copulas," in *Advances in Probability Distributions with Given Marginals* (G. Dall'Aglio, S. Kotz and G. Salinetti, eds.), pp. 13-50, Kluwer Academic Publishers, Dordrecht.

[29] Schweizer, B. and Sklar, A. (1983), *Probabilistic Metric Spaces*, Elsevier, New York.

[30] Schweizer, B. and Wolff, E. F. (1981), "On nonparametric measures of dependence for random variables," *Ann. Statist.* **9**, 870-885.

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