Editorial to the special issue on copulas of the Journal of the French Statistical Society

Titre: Éditorial du numéro spécial du journal de la SFdS sur les copules

Ivan Kojadinovic[1]

Copulas have been the subject of many (many many[2]) papers and have been applied in a wide variety of areas such as quantitative risk management [McNeil et al., 2005], econometric modeling [Patton, 2012], environmental modeling [Salvadori et al., 2007], to name a very few. The reason for what could be inelegantly called "the copula craze" lies in Sklar's theorem [Sklar, 1959] which states that every multivariate cumulative distribution function (c.d.f.) can be obtained by "coupling" together marginal c.d.f.s by means of a *copula*. From the point of view of the estimation of a multivariate distribution from data, this offers a great deal of flexibility as it allows practitioners to model the marginal c.d.f.s separately from the dependence structure, that is, the *copula*.

This editorial is organized as follows. In the first section, we define copulas and state Sklar's theorem. For those new to the subject, a short reading list is suggested in the second section. The third section presents the contents of this special issue.

1. Copulas and Sklar's theorem

What are copulas? Copulas are merely particular multivariate c.d.f.s. Let d be an integer greater than or equal to two.

Definition 1. *A d-dimensional copula is a c.d.f. on $[0,1]^d$ with standard uniform margins.*

The requirement that the marginal c.d.f.s be standard uniform is somehow arbitrary as discussed for instance by Embrechts in [Embrechts, 2009] who mentions that Hoeffding considered multivariate c.d.f.s whose margins are uniform on the interval $[-1/2, 1/2]$. In multivariate extreme-value theory, it is for instance more natural to consider multivariate c.d.f.s whose margins are unit Fréchet. The important message is that the way a multivariate distribution is "standardized" from the point of view of its margins does not alter the philosophy behind the concept of copula.

The key result at the origin of the increasing use of copulas for modeling multivariate distributions is the following theorem due to Sklar [Sklar, 1959].

---

[1] Laboratoire de mathématiques et applications, UMR CNRS 5142, Université de Pau et des Pays de l'Adour, B.P. 1155, 64013 Pau Cedex, France. E-mail: `ivan.kojadinovic@univ-pau.fr`

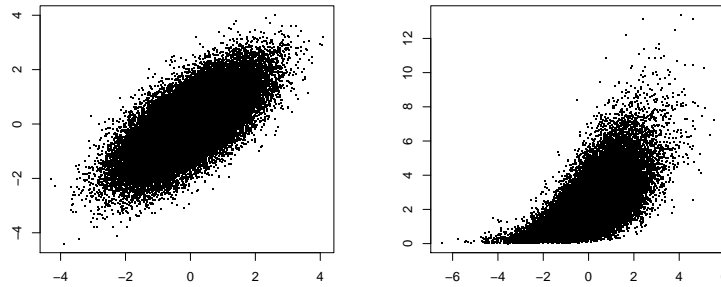[2] No, really, many many many. See [Genest et al., 2009a] for a bibliometric overview in finance.

FIGURE 1. *Fifty thousand independent realizations from two bivariate distributions constructed using* (1). *In both cases, C is the normal copula with parameter 0.7. The left plot was obtained by taking standard normal margins (the resulting distribution thus being the bivariate standard normal with correlation 0.7), while, in the right plot, a t and a gamma margin were used, respectively. The characteristics of the two underlying distributions are the same in terms of dependence as they are constructed from the same copula.*

Theorem 1. *Let F be a d-dimensional c.d.f. with marginal c.d.f.s $F_1, \ldots, F_d$. Then, there exists a copula $C : [0,1]^d \to [0,1]$ such that*

$$F(x_1, \ldots, x_d) = C\{F_1(x_1), \ldots, F_d(x_d)\}, \qquad \forall(x_1, \ldots, x_d) \in [-\infty, \infty]^d. \tag{1}$$

If the marginal c.d.f.s $F_1, \ldots, F_d$ are continuous, C is unique. Otherwise, C is uniquely determined on $\operatorname{Ran} F_1 \times \cdots \times \operatorname{Ran} F_d$, where, for any $i \in \{1, \ldots, d\}$, $\operatorname{Ran} F_i = F_i([-\infty, \infty])$ is the range of F_i. Conversely, if C is a copula and $F_1, \ldots, F_d$ are univariate c.d.f.s, then F defined in (1) is a d-dimensional c.d.f. whose marginal c.d.f.s are $F_1, \ldots, F_d$.

The name *copula* was introduced by Sklar as a reference to the fact that the function C in (1) "couples" together the marginal c.d.f.s $F_1, \ldots, F_d$ to form the multivariate c.d.f. F. A very interesting historical introduction to copulas is given by Durante and Sempi in [Durante and Sempi, 2010, Section 1.1] where the work of Sklar is presented as a continuation of that of Fréchet.

Given a univariate c.d.f. G, let G^{-1} denote its (left-continuous) generalized inverse, that is, $G^{-1}(u) = \inf\{x \in \mathbb{R} : G(x) \geq u\}$, $u \in [0, 1]$. Starting from (1) and under the assumption that the marginal c.d.f.s $F_1, \ldots, F_d$ are continuous, there holds

$$C(u_1, \ldots, u_d) = F\{F_1^{-1}(u_1), \ldots, F_d^{-1}(u_d)\}, \qquad \forall(u_1, \ldots, u_d) \in [0, 1]^d. \tag{2}$$

The above equation shows how copulas can be constructed from multivariate distributions with continuous margins. A typical example of this construction is the normal copula which is obtained by taking F in (2) equal to the multivariate standard normal c.d.f. An illustration of Sklar's theorem with the normal copula is given in Figure 1.

In typical applications of copulas, it is assumed that the multivariate distribution to be modeled has continuous margins. Sklar's theorem then implies that the underlying unknown copula is unique, which justifies its estimation from available data. Once the margins and the copula

are estimated, they are typically "glued" together as in (1), thereby providing the estimated multivariate c.d.f.

2. A very short and personal reading list

A smooth and statistically oriented introduction to copulas is given by Genest and Favre [Genest and Favre, 2007] (see also [Genest et al., 2009b] for additional material on goodness-of-fit testing). A "computational" complement to the previous paper is [Kojadinovic and Yan, 2010] where two real data sets are analyzed in R [R Development Core Team, 2013] using the copula package [Hofert et al., 2013].

A nice overview of the theory and practice with financial applications in mind is given by Embrechts [Embrechts, 2009]. An important related paper famous for warning financial data analysts about fallacies related to the use of linear correlation is [Embrechts et al., 2002].

An interesting historical perspective can be found in Durante and Sempi [Durante and Sempi, 2010]. The latter article and Chapter 5 of McNeil, Frey and Embrechts [McNeil et al., 2005] provide more detailed introductions.

3. Contents of the special issue

This special issue on copulas of the Journal of the French Statistical Society consists of ten articles, most of them of a methodological nature.

The first article, by Genest, Carabarín-Aguirre and Harvey, revisits the problem of copula parameter estimation: under the assumption that the unknown copula belongs to a given parametric family, how should the unknown parameter be estimated from the available data? The authors study, both theoretically and empirically, the properties of a method-of-moments estimator based on Blomqvist's beta. A similar issue is investigated by Hofert, Mächler and McNeil but from a more computational perspective. With financial applications in mind, the authors focus on the numerical challenges arising from the use and estimation of Archimedean copulas in high dimensions.

Estimating the parameter of the chosen copula family is a key step when modeling distributions with continuous margins. A very important related issue is whether the chosen family is a plausible choice given the data at hand. In the third article, Genest, Huang and Dufour revisit the goodness-of-fit problem and propose a test that is particularly sensitive to the fit in the tails of the copula. Quessy and Bellerive then investigate goodness-of-fit tests for the class of elliptical copulas by exploiting the particular structure of these models. Wang and Yan complement the work of Quessy and Bellerive by drawing attention to some practical issues arising when modeling multivariate distributions by means of elliptical copulas, and investigate conditional sampling algorithms for these distributions.

The problem of nonparametric estimation of a copula is the subject of the paper of Berghaus, Bücher and Dette. Focusing on the class of extreme value copulas, the authors study minimum distance estimators of the unknown Pickands dependence function (which characterizes the copula in this context) and propose related tests of extreme value dependence. The seventh article, by Ribatet and Sedki, puts forward the strong links existing between extreme value copulas and max-stable processes. The latter can be seen as infinite-dimensional extensions of multivariate

extreme value distributions and are increasingly used for environmental modeling. Also with environmental applications in mind, Salvadori, Durante and Perrone, propose, in the bivariate case, a method for constructing a piecewise-linear approximation of Kendall's distribution function which plays a key role in a recent definition of the notion of multivariate return period.

Typically, none of the usual textbook copula families fits data well in high dimensions. The concept of vine copula, increasingly used in finance and insurance, aims at improving this situation. In the eighth article, Czado, Jeske and Hofmann focus on the class of regular vine copulas, discuss their estimation and propose several selection strategies for choosing the underlying structure. Another class of copulas that was proposed with higher dimensional data in mind is that of hierarchical Kendall copulas. In the last article, Brechmann suggests several sampling algorithms for this class and evaluates them in a simulation study.

Acknowledgments

References

[Durante and Sempi, 2010] Durante, F. and Sempi, C. (2010). Copula theory: An introduction. In Jaworski, P., Durante, F., Härdle, W., and Rychlik, W., editors, *Copula theory and its applications (Warsaw, 2009)*, Lecture Notes in Statistics, pages 3–32. Springer-Verlag.

[Embrechts, 2009] Embrechts, P. (2009). Copulas: A personal view. *Journal of Risk and Insurance*, 76:639–650.

[Embrechts et al., 2002] Embrechts, P., McNeil, A., and Straumann, D. (2002). Correlation and dependence in risk management: Properties and pitfalls. In Dempster, M., editor, *Risk management: Value at risk and beyond*, pages 176–223. Cambridge University Press.

[Genest and Favre, 2007] Genest, C. and Favre, A.-C. (2007). Everything you always wanted to know about copula modeling but were afraid to ask. *Journal of Hydrological Engineering*, 12:347–368.

[Genest et al., 2009a] Genest, C., Gendron, M., and Bourdeau-Brien, M. (2009a). The advent of copulas in finance. *European Journal of Finance*, 15:609–618.

[Genest et al., 2009b] Genest, C., Rémillard, B., and Beaudoin, D. (2009b). Goodness-of-fit tests for copulas: A review and a power study. *Insurance: Mathematics and Economics*, 44:199–213.

[Hofert et al., 2013] Hofert, M., Kojadinovic, I., Mächler, M., and Yan, J. (2013). *copula: Multivariate dependence with copulas*. R package version 0.999-7.

[Kojadinovic and Yan, 2010] Kojadinovic, I. and Yan, J. (2010). Modeling multivariate distributions with continuous margins using the copula R package. *Journal of Statistical Software*, 34(9):1–20.

[McNeil et al., 2005] McNeil, A., Frey, R., and Embrechts, P. (2005). *Quantitative risk management*. Princeton University Press, New Jersey.

[Patton, 2012] Patton, A. (2012). Copula methods for forecasting multivariate time series. In *Handbook of Economic Forecasting*, volume 2. Springer Verlag.

[R Development Core Team, 2013] R Development Core Team (2013). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0.

[Salvadori et al., 2007] Salvadori, G., Michele, C. D., Kottegoda, N., and Rosso, R. (2007). *Extremes in Nature: An Approach Using Copulas*. Water Science and Technology Library, Vol. 56. Springer.

[Sklar, 1959] Sklar, A. (1959). Fonctions de répartition à *n* dimensions et leurs marges. *Publications de l'Institut de Statistique de l'Université de Paris*, 8:229–231.