Journal of Multivariate Analysis 140 (2015) 363–376

Contents lists available at ScienceDirect

Journal of Multivariate Analysis

journal homepage: www.elsevier.com/locate/jmva

A class of multivariate copulas based on products of bivariate copulas

Gildas Mazo *, Stéphane Girard, Florence Forbes

Inria and Laboratoire Jean Kuntzmann, Grenoble, France

ARTICLE INFO

Article history:
Received 8 July 2014
Available online 11 June 2015

AMS subject classifications:
62H99

Keywords:
Maximum-likelihood inference
Graphical models
Message-passing algorithm
Multivariate
Copula

ABSTRACT

Copulas are a useful tool to model multivariate distributions. While there exist various families of bivariate copulas, much less work has been done when the dimension is higher. We propose a class of multivariate copulas based on products of transformed bivariate copulas. The analytical forms of the copulas within this class allow to naturally associate a graphical structure which helps to visualize the dependencies and to compute the full joint likelihood even in high dimension. Numerical experiments are conducted both on simulated and real data thanks to a dedicated R package.

1. Introduction

The modeling of random multivariate events is a central problem in various scientific domains and the construction of multivariate distributions able to properly model the variables at play is challenging. A useful tool to deal with this problem is the concept of copula. Let $(X_1, \ldots, X_d)$ be a random vector with distribution function F. Let F_i be the (continuous) marginal distribution function of X_i, $i = 1, \ldots, d$. From Sklar's Theorem [27], there exists a unique function C such that

$$F(x_1, \ldots, x_d) = C(F_1(x_1), \ldots, F_d(x_d)), \quad (x_1, \ldots, x_d) \in \mathbb{R}^d. \tag{1}$$

This function C is called the copula of F and is the d-dimensional distribution function of the random vector $(F_1(X_1), \ldots, F_d(X_d))$. For a general account on copulas, see, e.g. [24]. Copulas are interesting since they permit to impose a dependence structure on pre-determined marginal distributions. While there exist many copulas in the bivariate case, it is less clear how to construct copulas in higher dimension. In the presence of non-Gaussianity and/or tail dependence, various constructions have been adopted, such as, for instance, Archimedean copulas [13], Vines [1] or elliptical copulas [5].

Archimedean copulas write

$$C(u_1, \ldots, u_d) = \psi(\psi^{-1}(u_1) + \cdots + \psi^{-1}(u_d)),$$

where ψ is a function from $[0, \infty)$ to $[0, 1]$ which has to verify certain properties for the copula to be well defined, see [23]. The generator ψ may be chosen in a given parametric family of functions. For instance, $\psi_\theta(t) = \exp(-t^{1/\theta})$, $\theta \geq 1$ yields the Gumbel family of copulas, see Example 1 in Section 3. Since there is a single parameter to model a d-dimensional phenomenon, this model is recognized not to be very flexible. Indeed, Archimedean copulas are exchangeable

* Corresponding author.
E-mail address: gildas.mazo@free.fr (G. Mazo).

http://dx.doi.org/10.1016/j.jmva.2015.06.001
0047-259X/

i.e. $C(u_1, \dots, u_d) = C(u_{\pi(1)}, \dots, u_{\pi(d)})$ for any permutation π of $\{1, \dots, d\}$. In particular, all pairs of variables share the same statistical distribution. These properties may not be desirable in practice.

Vines, on the opposite, achieve greater flexibility but at the price of increased complexity. As an illustration, we briefly describe a canonical vine copula – one of the two main types of vine copula models – through a decomposition of its density [1]:

$$c(u_1, \dots, u_d) = \prod_{j=1}^{d-1} \prod_{i=1}^{d-j} c_{j,j+1|1,\dots,j-1}\left(F(u_j|u_1, \dots, u_{j-1}), F(u_{j+i}|u_1, \dots, u_{j-1})\right)$$

where $c_{j,j+1|1,\dots,j-1}(\cdot, \cdot)$ is the (conditional) bivariate density of the jth and $(j+1)$th variables and where $F(\cdot|\cdot)$ represents the conditional distribution of the variables at play. When $d = 10$, there are more than one million possible decompositions, and, for each decomposition, there are many choices of parametric families for each conditional bivariate density in the product.

A third class of copulas to be presented in this introduction is the class of elliptical copulas. An elliptical copula is the copula of an elliptical distribution, whose density is given by [5,22]

$$f(x) = |\Sigma|^{-1/2} g\left((x-\mu)^\top \Sigma^{-1}(x-\mu)\right), \quad x \in \mathbb{R}^d,$$

for some positive definite matrix Σ and vector μ. The function g is called the density generator. This model implies, in particular, that if X has density f as above, then $X - \mu$ is distributed as $\mu - X$. This, in turn, implies that the lower and upper tail dependence coefficients (defined in Section 3) are equal, which is unrealistic in some applications, as, for example, extreme-value statistics. Moreover, elliptical copulas have in general as many as $O(d^2)$ parameters and it is thus difficult to carry out maximum likelihood inference [3] when d is large.

The main contribution of this paper is to propose a new class of multivariate copulas based on a product of bivariate copulas. The product is performed following the edges of a graph which permits to visualize the dependencies and to efficiently compute the likelihood, even in high dimension. The use of bivariate copulas as building blocks allows to take profit of the numerous parametric families proposed in the copula literature.

The rest of this paper is organized as follows. The new copula model is introduced in Section 2. Some links with Liebscher's construction [19] are stressed. Section 3 discusses some properties of the new copulas. The ability to construct new extreme-value models is highlighted. The dependence properties of bivariate marginals of the proposed class are also established. More specifically, some bounds are given on the most popular dependence coefficients (Spearman's rho and Kendall's tau) and on tail dependence coefficients. Section 4 is dedicated to the numerical aspects. A simulation procedure is provided and estimation by maximization of the pseudo-likelihood is discussed. The proposed copula model is applied in Section 5 to simulated and real datasets. Appendix gathers some proofs and technical details about the estimation procedure.

2. Constructing high dimensional copulas by multiplying bivariate ones

In this section, we propose a way to build high-dimensional copulas starting from bivariate ones. This construction allows one to take advantage of the large number of bivariate copulas introduced in the statistical literature. It is well known that a product of copulas is not a copula in general, the margins being no longer uniform. Roughly speaking, the new copula is thus obtained by multiplying bivariate copulas after a suitable transformation of the margins. The main feature of the new copula is that it can be associated with a graph describing the dependencies between the variables. To be more specific, let $U_1, \dots, U_d$ be d standard uniform random variables and denote by $\{ij\}$ the index of the pair (U_i, U_j). Introduce $E \subset \{\{ij\} : i, j = 1, \dots, d, \ j > i\}$ a subset of the set of all pair indices. The cardinal of E, denoted by $|E|$, is less or equal to $d(d-1)/2$. The pair index $e \in E$ is said to contain the variable index i if there exists $k \neq i$ such that $e = \{ik\}$ or $e = \{ki\}$. For all $i = 1, \dots, d$, let $N(i)$ be the set of neighbors of i defined as $N(i) = \{e \in E \text{ such that } e \text{ contains } i\}$ and introduce $n_i := |N(i)|$. It is then natural to associate a graph to the set E as follows: an element $e = \{ij\} \in E$ is an edge linking U_i and U_j in the graph whose nodes are the variables $U_1, \dots, U_d$. The example $E = \{\{12\}, \{24\}, \{23\}, \{35\}\}$ is illustrated in Fig. 1. For $u = (u_1, \dots, u_d) \in [0, 1]^d$, consider the functional

$$C(u_1, \dots, u_d) = \prod_{\{ij\} \in E} \tilde{C}_{ij}\left(u_i^{1/n_i}, u_j^{1/n_j}\right), \qquad (2)$$

where the $\tilde{C}_{ij}$'s are arbitrary bivariate copulas for all $\{ij\} \in E$. Keeping in mind the graphical representation associated with E, the function C defined in (2) is a product over the edges of the graph. For instance, when $E = \{\{12\}, \{24\}, \{23\}, \{35\}\}$ as in Fig. 1, function (2) can be written as

$$C(u_1, u_2, u_3, u_4, u_5) = \tilde{C}_{12}\left(u_1, u_2^{1/3}\right) \tilde{C}_{24}\left(u_2^{1/3}, u_4\right) \tilde{C}_{23}\left(u_2^{1/3}, u_3^{1/2}\right) \tilde{C}_{35}\left(u_3^{1/2}, u_5\right).$$

In the following, (2) is referred to as the Product of Bivariate Copulas (PBC) copula, or PBC model. The next result establishes that (2) is a copula.

Proposition 1. *PBC (2) is a well defined copula.*