Journal of Multivariate Analysis 110 (2012) 74–90

Contents lists available at SciVerse ScienceDirect

Journal of Multivariate Analysis

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

Beyond simplified pair-copula constructions

Elif F. Acar *, Christian Genest, Johanna Nešlehová

Department of Mathematics and Statistics, McGill University, Montréal, QC, Canada H3A 0B9

ARTICLE INFO

Article history:
Available online 17 February 2012

AMS subject classifications:
62F12
62H12
62H20

Keywords:
Conditional copulas
Kendall's tau
Local likelihood
Pair-copula constructions
Ranks
Vines

ABSTRACT

Pair-copula constructions (PCCs) offer great flexibility in modeling multivariate dependence. For inference purposes, however, conditional pair-copulas are often assumed to depend on the conditioning variables only indirectly through the conditional margins. The authors show here that this assumption can be misleading. To assess its validity in trivariate PCCs, they propose a visual tool based on a local likelihood estimator of the conditional copula parameter which does not rely on the simplifying assumption. They establish the consistency of the estimator and assess its performance in finite samples via Monte Carlo simulations. They also provide a real data application.

1. Introduction

Over the past two decades, dependence modeling via copulas has evolved considerably and has found applications in areas as diverse as actuarial science, biostatistics, finance, hydrology, and machine learning. In the bivariate case, many parametric copula families have been proposed that can represent a broad range of dependence patterns. In higher dimensions, however, parametric copula families are harder to construct and their tractability often comes at the cost of flexibility. For example, meta-elliptical copulas are somewhat of a straightjacket, if only because all lower-dimensional margins belong to the same class. In many applications, this property is too restrictive as pairs of variables may exhibit very different dependence patterns.

A more flexible way to model multivariate dependences is offered by pair-copula constructions (PCCs), also known as *vine copulas* [7,15,16]. Vines are graphical models that provide a systematic way to decompose a multivariate copula into a cascade of bivariate copulas, some of which are conditional. A simple example of a PCC in the trivariate case consists of writing the joint density c of a random vector (U_1, U_2, U_3) with uniform margins on $(0, 1)$ in the form

$$c(u_1, u_2, u_3) = c_{12}(u_1, u_2)c_{23}(u_2, u_3)c_{13|2}(u_{1|2}, u_{3|2}; u_2). \quad (1)$$

Here, c_{12} and c_{23} are the copula densities of the pairs (U_1, U_2) and (U_2, U_3), respectively. Furthermore, $c_{13|2}$ is the conditional copula density of the pair (U_1, U_3) given $U_2 = u_2$, evaluated at $u_{k|2} = \Pr(U_k \le u_k|U_2 = u_2)$ for $k = 1, 3$. Any choice of c_{12}, c_{13} and $c_{13|2}$ leads to a valid trivariate copula density. More generally, using different bivariate copulas as building blocks in a d-variate PCC, one can construct highly flexible multivariate copula models.

Inference for a given PCC is typically carried out by specifying a parametric copula for each building block. Copula parameters are then estimated sequentially starting with the unconditional pair-copulas and moving up the hierarchy [1].

* Corresponding author.
E-mail address: eacar@math.mcgill.ca (E.F. Acar).

0047-259X/$ – see front matter © 2012 Elsevier Inc. All rights reserved.
doi:10.1016/j.jmva.2012.02.001

In the above example, this amounts to estimating the parameters of c_{12} and c_{23} first, and those of $c_{13|2}$ in the second step. A standard assumption is that the conditional pair-copulas of the PCC depend on the conditioning variable(s) only through the conditional margins. In (1), this is equivalent to assuming that the conditional copula $c_{13|2}$ of the pair (U_1, U_3) given $U_2 = u_2$ is *the same* for all values of $u_2 \in (0, 1)$.

This simplifying assumption seems to have been made mainly for convenience at a time when inference tools for conditional copulas were still under development [20]. Through examples, it is shown in [14] that simplified PCCs can provide a good approximation in some cases. This paper revisits this issue and introduces a nonparametric smoothing methodology that relaxes this simplifying assumption for trivariate PCCs.

After a brief summary of vine copula constructions in the trivariate case in Section 2, estimation for simplified three-dimensional PCCs is described in Section 3. Through simulations, it is then shown in Section 4 that inference based on simplified PCCs can be misleading and may even conduce the belief that some pairs of variables are conditionally independent when in fact they are not. The new methodology, which derives from recent work [4], is described in Section 5. This approach is seen to perform well in simulations and in a data application, as detailed in Sections 6 and 8, respectively. The consistency of the proposed method is presented in Section 7 and Section 9 concludes with a short discussion.

The following notation is used throughout the paper. Vectors in $\mathbb{R}^3$ are denoted by bold letters, e.g., $\mathbf{x} = (x_1, x_2, x_3) \in \mathbb{R}^3$. If $A \subset \{1, 2, 3\}$ is non-empty, $\mathbf{x}_A$ stands for an $|A|$-dimensional vector with components x_k, $k \in A$. If $\mathbf{X}$ is a random vector with distribution function F and density f, then for arbitrary disjoint index sets A and B, the symbols $F_{A|B}$ and $f_{A|B}$ denote the conditional distribution function and density of $\mathbf{X}_A$ given $\mathbf{X}_B = \mathbf{x}_B$, respectively.

2. Trivariate PCCs

Let X_1, X_2, X_3 be random variables with joint distribution function F and continuous margins F_1, F_2, F_3, respectively. Sklar's Representation Theorem [22] states that, for all $x_1, x_2, x_3 \in \mathbb{R}$,

$$F(x_1, x_2, x_3) = C\{F_1(x_1), F_2(x_2), F_3(x_3)\},$$

where C is a copula, i.e., a distribution function with margins that are uniform on $(0, 1)$. If F is absolutely continuous, its density can be written in terms of the density c of C as

$$f(x_1, x_2, x_3) = c\{F_1(x_1), F_2(x_2), F_3(x_3)\} \prod_{k=1}^{3} f_k(x_k),$$

where, for each $k \in \{1, 2, 3\}$, f_k is the density of F_k.

A PCC is based on the fact that f can be decomposed as

$$f(x_1, x_2, x_3) = f_3(x_3) \times f_{2|3}(x_2|x_3) \times f_{1|23}(x_1|x_2, x_3). \tag{2}$$

Note that this factorization is unique up to relabeling. For any index set $A \subset \{1, 2, 3\}$ and $k \in A$, let $A - k = A \setminus \{k\}$. Using Sklar's Representation Theorem, one can then write, for arbitrary $j \notin A$,

$$f_{j|A} = c_{jk|A-k}(F_{j|A-k}, F_{k|A-k}) f_{j|A-k}. \tag{3}$$

Repeated applications of relation (3) in (2) make it possible to express f as

$$\begin{aligned} f(x_1, x_2, x_3) &= f_1(x_1) f_2(x_2) f_3(x_3) \times c_{12}\{F_1(x_1), F_2(x_2)\} \times c_{23}\{F_2(x_2), F_3(x_3)\} \\ &\quad \times c_{13|2}\{F_{1|2}(x_1|x_2), F_{3|2}(x_3|x_2); x_2\}, \end{aligned} \tag{4}$$

which reduces to (1) if the margins of F are uniform. The univariate conditional distributions featuring in (4) are given by

$$F_{j|k}(x_j|x_k) = h_{jk}\{F_j(x_j), F_k(x_k)\},$$

where, for all $u, v \in (0, 1)$,

$$h_{jk}(u, v) = \frac{\partial}{\partial v} C_{jk}(u, v). \tag{5}$$

3. Inference for simplified PCCs

Now suppose the density f of (X_1, X_2, X_3) follows a simplified PCC model, i.e., f is of the form (4), where the conditional copula density $c_{13|2}$ does not depend on the conditioning variable. The last term in (4) thus reduces to

$$c_{13|2}\{F_{1|2}(x_1|x_2), F_{3|2}(x_3|x_2)\}.$$

To ease the presentation, assume that all copulas appearing in (4) are parametrized by scalar parameters $\theta_{12}, \theta_{23}, \theta_{13|2}$ that are indexed in the same way as the corresponding copula.