Review of Business and Economic Literature
July-December (2018)
ISSN 2031-1761

An Application of Kendall Distributions

by U. dos Anjos and N. Kolev

Ulisses dos Anjos
University of São Paulo,
Department of Statistics, Brazil

Nikolai Kolev
University of São Paulo,
Department of Statistics, Brazil

ABSTRACT

We propose to use copulas in order to describe the dependence structure between nonoverlapping random vectors with known multivariate marginals scanned by their Kendall distribution functions.

I. INTRODUCTION

The dependence between real-valued random variables $X_1, \ldots, X_n$ is completely described by their joint distribution function $H(x_1, \ldots, x_n)$. Necessary and sufficient conditions for a right-continuous function H: $[-\infty,\infty]^n \mapsto [0,1]$ to be a multivariate distribution function are:

(i) $\lim_{x_j \to -\infty} H(x_1, \ldots, x_n) = 0$ for any $j = 1, \ldots, n$;

(ii) $\lim_{x_j \to \infty} H(x_1, \ldots, x_n) = 1$ for all $j = 1, \ldots, n$ and

(iii) for all $(a_1, \ldots, a_n)$, $(b_1, \ldots, b_n)$ with $a_j < b_j$,

$$\sum_{i_1=1}^{2} \cdots \sum_{i_n=1}^{2} (-1)^{i_1+\ldots+i_n} H(x_{1i_1}, \ldots, x_{ni_n}) \geq 0, \tag{1}$$

where $x_{j1} = a_j$ and $x_{j2} = b_j$, $j = 1, \ldots, n$. Note that if H has nth-order derivatives, then condition (1) is equivalent to $\frac{\partial_n H}{\partial_{x_1} \ldots \partial_{x_n}} \geq 0$.

The problem of constructing multivariate distributions with given univariate marginals has aroused considerable interest. For example, the idea of separating H into a part which describes the dependence structure and parts which describe the marginal behavior only, has lead to the concept of copula, see Nelsen (1999).

Assume that the continuous random variables $X_1, \ldots, X_n$ have known marginal distributions $F_{X_1}, \ldots, F_{X_n}$, correspondingly. According to Sklar's theorem there exist unique copula C_H such that

$$H(x_1, \ldots, x_n) = C_H(F_{X_1}(x_1), \ldots, F_{X_n}(x_n)) \tag{2}$$

for all $(x_1, \ldots, x_n) \in [-\infty, \infty]^n$.

Sklar's theorem holds whenever the dimension $n \geq 2$, so most of the results could be used. But note that since it is much more convenient to work in dimension $n = 2$, practitioners usually wish to aggregate two-dimensional framework to obtain a multidimensional one.

Even though the construction of families of joint distribution functions for given univariate marginals has been so widely studied, the case of higher dimensional marginals has been focused more on study of compatibility of overlapping marginals and bounds for the corresponding Fréchet classes, see Joe (1997), Chapter 3.

Nevertheless, one cannot just select a parametric family of functions with the right boundary properties and expect them to satisfy the rectangle condition (1) of a multivariate joint distribution function.

Generally, a family of multivariate distributions must be constructed through methods such as mixtures, stochastic representations and limits, see Joe (1997), Chapter 4.

The insufficiency of the copula function to handle multivariate distributions with given marginals is illustrated by the Nutshell copula's paradox discussed by Genest *et al.* (1995). They showed that the only possibility that

$$H(x_1, \ldots, x_{n_1}, x_{n_1+1}, \ldots, x_{n_1+n_2})$$
$$= C(H_{n_1}(x_1, \ldots, x_{n_1}), H_{n_2}(x_{n_1+1}, \ldots, x_{n_1+n_2}))$$

defines a (n_1+n_2)-dimensional distribution function, $n_1+n_2 \geq 3$, for all H_{n_1} and H_{n_2} (with dimensions n_1 and n_2, respectively) is $C(u,v) = uv$, i.e. the independence copula.

A well known and difficult problem of multivariate distribution theory is to construct a multivariate distribution H with prescribed multivariate marginals. Some aspects of this problem are discussed by Dall'Aglio (1972). The problem of existence was solved by Kellerer (1964a) but the solution is essentially of theoretical kind and does not allow to use these results in most of the practical situations. The cases, which allow a simple construction are classified by Kellerer (1964b). A typical example is when a common distribution can be constructed as a distribution of a Markov chain.

Marco and Ruiz-Rivas (1992) are concerned with the following problem: Given d possibly multivariate nonoverlapping marginal distributions $H_1, \ldots, H_d$ of dimensions $n_1, \ldots, n_d$, respectively, what conditions should a d-dimensional function C_d satisfy in order for $C_d(H_1, \ldots, H_d)$ to be a $(\sum_{i=1}^{d} n_i)$-dimensional distribution function. They also give a procedure for construction of such a function C_d.

Cuadras (1992) provides a method of constructing multivariate distributions where both the multivariate marginals and inter-correlation matrix are given.

Li *et al.* (1996) introduced a new tool, called linkage (based on conditional distributions), which is useful for handling multivariate distributions with given multivariate marginals. In Li *et al.* (1999) the authors extend the linkage function to the dynamic linkage, which can usefully model multivariate distributions of non-negative random variables by taking advantage of the time dynamics of the underlying life times.

Our approach presented in Section II is completely different than the above research. The purpose is to provide a simple method, based on Kendall distribution functions, recently introduced by Nelsen *et al.* (2003), in order to represent the dependence structure between non-overlapping marginals which are in general multivariate. The method presented is motivated by the following fact. Typically, one has an idea about the dependence mechanism of nonoverlapping segments composing a given portfolio, and the basic interest is to search a dependence structure between those segments.

II. COPULAS WITH KENDALL MARGINAL DISTRIBUTIONS

Let us suppose that the continuous random vector $(X_1, \dots, X_n)$ is partitioned into d nonoverlapping clusters, $2 \le d \le n$, and let the l-th cluster contains n_l variables $(X_{m_{l-1}+1}, \dots, X_{m_l})$ with $m_l = \sum_{j=1}^{l} n_j$, $m_0 = 0$, $l = 1, \dots, d$, and $n = n_1 + \dots + n_d$. We presume that the clusters can even be independent.

Assume that we know the within cluster dependence given by the marginal distributions H_l and their associated copulas C_{H_l} for $l = 1, \dots, d$. Consider the sequence $U_1, \dots, U_n$ of standard uniformly distributed random variables, i.e. $F_{X_i}(X_i) = U_i \sim U(0,1)$, $i = 1, \dots, n$. Let

$$\Psi_l = H_l(X_{m_{l-1}+1}, \dots, X_{m_l}) = \mathrm{C}_{\mathrm{H}_1}(U_{m_{l-1}+1}, \dots, U_{m_l})$$

be the random variable with distribution function

$$K_l(\psi_l) = P(\Psi_l \le \psi_l) = P(C_{H_l}(U_{m_{l-1}+1}, \dots, U_{m_l}) \le \psi_l), \quad \psi_l \in [0,1],$$

which, in fact, is the Kendall distribution function of $(X_{m_{l-1}+1}, \dots, X_{m_l})$, see Nelsen *et al.* (2003). Hence, $W_l = K_l(\Psi_l) \sim U(0,1)$, $l = 1, \dots, d$.

The function K_l depends only on the copula C_{H_l} associated to H_l and therefore, K_l is a one dimensional summary of the dependence given by C_{H_l}. Note that if $n_l = 1$ for some $l = 1, \dots, d$, we have

$$K_l(\psi_l) = P(\Psi_l \le \psi_l) = P(U_{m_{l-1}+1} \le \psi_l) = \psi_l.$$

Since K_l are continuous, according to Sklar's theorem there exist an unique d-dimensional copula C_d which is a distribution function of the vector $(W_1, \dots, W_d)$, i.e.

$$C_d(w_1, \dots, w_d) = P(W_1 \le w_1, \dots, W_d \le w_d) \tag{3}$$

for $(w_1, \dots, w_d) \in [0,1]^d$.

Let us underline that the function C_d describes just the dependence structure between the nonoverlapping clusters that we need. Relation (3) can be considered as a statement of a version of the Sklar's theorem for copulas with nonoverlapping multivariate marginals. If substitute $d = n$ in (3), one gets (2), i.e. the conclusion of the Sklar's theorem (when the univariate marginals are known or fixed).

For better understanding of this result we provide two examples. For both, consider the random vector (X_1, X_2, X_3, X_4), with given bivariate marginals $H_1(x_1, x_2) = P(X_1 \leq x_1, X_2 \leq x_2)$ and $H_2(x_3, x_4) = P(X_3 \leq x_3, X_4 \leq x_4)$, and let C_1 and C_2 be the associated copulas, correspondingly. Let G be the joint distribution function of the random variables $\Psi_1 = C_{H_1}(U_1, U_2)$ and $\Psi_2 = C_{H_2}(U_3, U_4)$, i.e. $G(\psi_1, \psi_2) = P(\Psi_1 \leq \psi_1, \Psi_2 \leq \psi_2)$ and C_G be the associated copula.

Example 1

Consider the copulas $C_{H_1}(u_1, u_2) = min(u_1, u_2)$ *and* $C_{H_2}(u_3, u_4) = min(u_3, u_4)$. *Let* $C_G(v_1, v_2) = v_1 v_2 + \theta v_1 v_2 (1 - v_1)(1 - v_2)$, *be the copula corresponding to the dependence structure between the clusters* (X_1, X_2) *and* (X_3, X_4).

Let K_1 *and* K_2 *be the Kendall distributions associated to* (X_1, X_2) *and* (X_3, X_4), *and therefore to the copulas* C_{H_1} *and* C_{H_2}, *respectively. In this case we have* $K_1(\psi_1) = \psi_1$ *and* $K_2(\psi_2) = \psi_2$, *e.g. Nelsen et al. (2001), where* $\psi_1 = C_{H_1}(u_1, u_2)$ *and* $\psi_2 = C_{H_2}(u_3, u_4)$. *Then, the joint distribution of the random vector* (Ψ_1, Ψ_2) *is given by*

$$G(\psi_1, \psi_2) = C_G(K_1(\psi_1), K_2(\psi_2)) = \psi_1 \psi_2 + \theta \psi_1 \psi_2 (1 - \psi_1)(1 - \psi_2),$$

or equivalently,

$$G(C_{H_1}(u_1, u_2), C_{H_2}(u_3, u_4)) = min(u_1, u_2) min(u_3, u_4) + \theta min(u_1, u_2) \times min(u_3, u_4)[1 - min(u_1, u_2)][1 - min(u_3, u_4)].$$

Example 2

Let $C_{H_1}(u_1, u_2) = min(u_1, u_2)$ *and* $C_{H_2}(u_3, u_4) = max(u_3 + u_4 - 1, 0)$. *Then, the Kendall distributions associated to the copulas* C_{H_1} *and* C_{H_2} *are* $K_1(\psi_1) = \psi_1$ *and* $K_2(\psi_2) = 1$, *see Nelsen et al. (2001). For the same* *copula* C_G *of Example 1 we obtain* $G(\psi_1, \psi_2) = C_G(K_1(\psi_1), K_2(\psi_2)) = \psi_1 + \theta\psi_1(1 - \psi_1)(1 - 1) = \psi_1$, *i.e.* $G(C_{H_1}(u_1, u_2), C_{H_2}(u_3, u_4)) = min(u_1, u_2)$.

As one can see, the Kendall distribution associated to the copula $C_{H_2}(u_3, u_4) = max(u_3 + u_4 - 1,0)$ in Example 2 is the degenerated one, so we lost all information about (U_3, U_4). Therefore, the joint distribution function $G(\psi_1, \psi_2)$ depends only on u_1 and u_2.

III. CONCLUSION

The need of models in specific situations when different vectors of data should be studied jointly is clear. The problems of this kind arise if one needs to build a stochastic model in a situation where the information about the kind of dependence and knowledge of certain marginal distributions is available. In this note we propose a new way to use the Sklar's theorem to represent the dependence structure between nonoverlapping random vectors via Kendall distribution function.

One needs to impose additional restrictions on marginals H_1 in order C_d given by (3) to be an n-dimensional distribution function of $(X_1, \dots, X_n)$. As Marco and Ruiz-Rivas (1992) showed, in the simplest case, for an arbitrary d-dimensional copula those restrictions lead to max-infinitely divisible marginals. The Nutshell copula's paradox commented by Genest *et al.* (1995) is a consequence of such limitations.

REFERENCES

Cuadras, C., 1992, Probability Distributions with Given Multivariate Marginals and Given Dependence Structure, *Journal of Multivariate Analysis* 42, 51-66.

Dall'Aglio, G., 1972, Fréchet Classes and Compatibility of Distribution Functions, *Symposia Mathematica* 9, 131-150.

Genest, C., Quesada-Molina, J.J., Rodrigues-Lallena, J.A., 1995, De l'impossibilité de construire des lois à marges multidimensionalles données à partir de copules, *Comptes Rendus de l'Académie des Sciences* 320, 723-726.

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

Kellerer, H., 1964a, Masstheoretische Marginalproblem, *Mathematische Annalen* 153, 168-198.

Kellerer, H., 1964b, Verteilungsfunctionen mit gegebenen Marginalverteilungen, *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete* 3, 247-270.

Li, H., Scarsini, M., Shaked, M., 1996, Linkages: a Tool for the Construction of Multivariate Distributions with Given Nonoverlapping Multivariate Marginals, *Journal of Multivariate Analysis* 56, 20-41.

Li, H., Scarsini, M., Shaked, M., 1999, Dynamic Linkages for Multivariate Distributions with Given Nonoverlapping Multivariate Marginals, *Journal of Multivariate Analysis* 68, 54-77.
Marco, J.M., Ruiz-Rivas, C., 1992, On the Construction of Multivariate Distributions with Given Nonoverlapping Multivariate Marginals, *Statistics & Probability Letters* 15, 259-265.
Nelsen, R., 1999, An Introduction to Copulas, (Springer, New York).
Nelsen, R., Quesada-Molina, J., Rodrigues-Lallena, J., and Ubeda-Flores, M., 2001, Distribution Function of Copulas: a Class of Bivariate Probability Integral Transforms, *Statistics & Probability Letters* 54, 277-282.
Nelsen, R., Quesada-Molina, J., Rodrigues-Lallena, J., Ubeda-Flores, M., 2003, Kendall Distribution Functions, *Statistics & Probability Letters* 65, 263-268.
Rüschendorf, L., 1985, Construction of Multivariate Distributions with Given Marginals, *Annals of the Institute of Statistical Mathematics* 37 part A, 225-233.
Schweizer, R., Sklar, E., 1981, Probabilistic Metric Spaces, (Elsevier, New York).