Brazilian Journal of Probability and Statistics (2005), 19, pp. 85–92.

A method for constructing multivariate distributions with given bivariate margins

Ali Dolati[1] and **Manuel Úbeda-Flores**[2]
[1] *Yazd University*
[2] *Universidad de Almería*

Abstract: In this paper we provide a method for constructing multivariate distributions with given bivariate margins and dependence structure. We also measure the dependence of any distribution constructed with this method and illustrate our results with examples.

Key words: Copulas, multivariate marginals, ordering, Spearman's rho.

1 Introduction

The construction of multivariate distributions with given margins has been a problem of interest to statisticians for many years: Nelsen (1999) summarizes different methods of constructing copulas (distributions with uniform univariate margins). A difficult problem related to the theory of multivariate distributions is to construct a multivariate distribution with prescribed multivariate margins. Some aspects of this problem including existence of such distributions, compatibility and methods of constructing are discussed, for instance, by Dall'Aglio (1972), Cohen (1984), Rüschendorf (1985), Cuadras (1992), Marco and Ruiz-Rivas (1992), Li *et al.* (1996,1999) and Joe (1997). Our approach presented in Section 2 is—in some sense—different from the above ones. We provide a method for constructing multivariate distributions with given bivariate margins and dependence structure. Problems of this kind arise if one needs to build a model in a situation where the information about the type of dependence structure of bivariate margins of a random vector is available and the basic interest is to search a dependence structure among all the components of that vector.

We now review the concept of a copula (for a complete study, see Nelsen, 1999). Let n be a natural number such that $n \geq 2$. An n-dimensional *copula* (briefly n-copula) is a function $C\colon [0,1]^n \longrightarrow [0,1]$ which satisfies:

(C1) For every $\mathbf{u} = (u_1, u_2, \ldots, u_n)$ in $[0,1]^n$, $C(\mathbf{u}) = 0$ if at least one coordinate of $\mathbf{u}$ is 0, and $C(\mathbf{u}) = u_k$ whenever all coordinates of $\mathbf{u}$ are 1 except u_k; and

(C2) for every $\mathbf{a} = (a_1, a_2, \ldots, a_n)$ and $\mathbf{b} = (b_1, b_2, \ldots, b_n)$ in $[0,1]^n$ such that $a_k \leq b_k$ for all $k = 1, 2, \ldots, n$, $V_C([\mathbf{a},\mathbf{b}]) = \sum \operatorname{sgn}(\mathbf{c}) C(\mathbf{c}) \geq 0$ where $[\mathbf{a},\mathbf{b}]$ denotes the *n-box* $[a_1, b_1] \times [a_2, b_2] \times \cdots \times [a_n, b_n]$, the sum is taken over all the *vertices* $\mathbf{c} = (c_1, c_2, \ldots, c_n)$ of $[\mathbf{a},\mathbf{b}]$ such that each c_k is equal to either a_k or b_k, and $\operatorname{sgn}(\mathbf{c})$

is 1 if $c_k = a_k$ for an even number of k's, and -1 if $c_k = a_k$ for an odd number of k's.

The importance of copulas as a tool for statistical analysis and modelling stems largely from the observation that the joint distribution H of a set of $n \geq 2$ random variables X_i with marginals F_i can be expressed by $H(\mathbf{x}) = C(F_1(x_1), F_2(x_2), \ldots, F_n(x_n))$, $\mathbf{x} = (x_1, x_2, \ldots, x_n) \in [-\infty, \infty]^n$, in terms of a copula C that is uniquely determined on $\text{Ran}F_1 \times \text{Ran}F_2 \times \cdots \times \text{Ran}F_n$. Let Π^n denote the copula of independent random variables, i.e., $\Pi^n(\mathbf{u}) = \prod_{i=1}^{n} u_i$; and let $\mathbf{U} > \mathbf{u}$ denote the point-wise inequality $(U_1 > u_1, U_2 > u_2, \ldots, U_n > u_n)$, where $\mathbf{U}$ is a random vector with n-copula C. The *survival function* associated to C is defined by $\overline{C}(\mathbf{u}) = P[\mathbf{U} > \mathbf{u}]$, i.e.,

$$\overline{C}(\mathbf{u}) = 1 + \sum_{k=1}^{n} (-1)^k \sum_{1 \leq i_1 < i_2 < \cdots < i_k \leq n} C_{i_1 i_2 \cdots i_k}(u_{i_1}, u_{i_2}, \ldots, u_{i_k}), \tag{1.1}$$

where the copulas on the right-hand side are appropriate lower dimensional margins. Finally, if C_1 and C_2 are two n-copulas, $C_1 \leq C_2$ denotes the inequality $C_1(\mathbf{u}) \leq C_2(\mathbf{u})$ for all $\mathbf{u}$ in $[0, 1]^n$.

2 Construction

In this section, we study a new function in order to provide a procedure for constructing families of n-copulas with given bivariate margins and dependence structure. We will also measure the dependence among n random variables whose associated n-copula is given by this construction in terms of the measures associated to its bivariate margins. The following theorem shows the main result of this paper.

Theorem 2.1 *Let $\{C_{ij} : 1 \leq i < j \leq n\}$ be a set of $\binom{n}{2}$ 2-copulas, and let $C\colon [0, 1]^n \longrightarrow [0, 1]$ be the function defined by*

$$C(\mathbf{u}) = \sum_{1 \leq i < j \leq n} C_{ij}(u_i, u_j) \prod_{\substack{k=1 \\ k \neq i,j}}^{n} u_k - \frac{(n-2)(n+1)}{2} \prod_{i=1}^{n} u_i. \tag{2.1}$$

Then C is an n-copula whose bivariate margins are C_{ij} if and only if

$$\sum_{1 \leq i < j \leq n} \frac{V_{C_{ij}}([u_i, v_i] \times [u_j, v_j])}{(v_j - u_j)(v_i - u_i)} \geq \frac{(n-2)(n+1)}{2} \tag{2.2}$$

for every u_k, v_k in $[0, 1]$, $k = 1, 2, \ldots, n$, such that $u_k < v_k$.

Proof. Let C be the function given by (2.1). For every $\mathbf{u}$ in $[0, 1]^n$, it is immediate that $C(u_1, \ldots, u_{i-1}, 0, u_{i+1}, \ldots, u_n) = 0$ and $C(1, \ldots, 1, u_i, 1, \ldots, 1) = u_i$ for all

$i = 1, 2, \dots, n$, whence condition (C1) is satisfied. To verify condition (C2), let $\mathbf{u}$ and $\mathbf{v}$ be two points in $[0,1]^n$ such that $u_k \le v_k$ for all $k = 1, 2, \dots, n$. Then, after some elementary algebra, we have that

$$V_C([u_1, v_1] \times [u_2, v_2] \times \cdots \times [u_n, v_n]) = \sum_{1 \le i < j \le n} V_{C_{ij}}([u_i, v_i] \times [u_j, v_j]) \prod_{\substack{k=1 \\ k \ne i,j}}^{n} (v_k - u_k) - \frac{(n-2)(n+1)}{2} \prod_{k=1}^{n} (v_k - u_k) \ge 0.$$

If there exists $i \in \{1, 2, \dots, n\}$ such that $u_i = v_i$, then the result is trivial; otherwise, the inequality (2.2) holds. Conversely, we only need to follow the same steps backwards. Finally, we have $C(1, \dots, 1, u_i, 1, \dots, 1, u_j, 1, \dots, 1) = C_{ij}(u_i, u_j)$, $1 \le i < j \le n$, which completes the proof. ■

Remark 2.1 It is easy to check that each $l-$margin, $3 \le l < n$, of C has also the form given by (2.1) for appropriate dimension l. Note also that if all the bivariate margins in Theorem 2.1 are absolutely continuous then the inequality (2.2) is equivalent to

$$\sum_{1 \le i < j \le n} c_{ij}(u_i, u_j) \ge \frac{(n-2)(n+1)}{2}, \tag{2.3}$$

where c_{ij} denotes the density function of each margin C_{ij}.

We now investigate a partial ordering and a positive dependence property on the n-copulas defined by (2.1). Given two n-copulas C_1 and C_2, C_1 is said *more concordant* than C_2 if $C_1(\mathbf{u}) \ge C_2(\mathbf{u})$ and $\overline{C}_1(\mathbf{u}) \ge \overline{C}_2(\mathbf{u})$ for all $\mathbf{u}$ in $[0,1]^n$. In the bivariate case, the two above inequalities are equivalent. When $C_2 = \Pi^n$, C_1 is said *positively orthant dependent* (POD); and, in the bivariate case, C_1 is said *positively quadrant dependent* (PQD). For more details, see Nelsen (1999). Before giving the result, we need a preliminary lemma whose proof is straightforward using (1.1).

Lemma 2.2 *Let C be an n-copula given by* (2.1) *via Theorem* 2.1*. Then*

$$\overline{C}(\mathbf{u}) = \prod_{i=1}^{n} (1 - u_i) + \sum_{1 \le i < j \le n} (C_{ij}(u_i, u_j) - u_i u_j) \prod_{\substack{k=1 \\ k \ne i,j}}^{n} (1 - u_k).$$

Theorem 2.3 *Let $\{C_{ij} : 1 \le i < j \le n\}$ and $\{C'_{ij} : 1 \le i < j \le n\}$ be two sets of $\binom{n}{2}$ margins* 2*-copulas of two n-copulas C and C', respectively, defined by* (2.1) *via Theorem* 2.1*, and such that $C_{ij} \ge C'_{ij}$, $1 \le i < j \le n$. Then C is more concordant than C'.*

Proof. Let C and C' be two n-copulas as in the hypothesis. Then, for every $\mathbf{u}$ in $[0,1]^n$, we have that $C(\mathbf{u}) \geq C'(\mathbf{u})$ if and only if

$$\sum_{1\leq i<j\leq n} [C_{ij}(u_i,u_j) - C'_{ij}(u_i,u_j)] \prod_{\substack{k=1\\k\neq i,j}}^{n} u_k \geq 0.$$

On the other hand, when $n \geq 3$—recall that if $n = 2$ we do not need to study this part—, using Lemma 2.2, we have that $\overline{C}(\mathbf{u}) \geq \overline{C'}(\mathbf{u})$ is equivalent to

$$\sum_{1\leq i<j\leq n} [C_{ij}(u_i,u_j) - C'_{ij}(u_i,u_j)] \prod_{\substack{k=1\\k\neq i,j}}^{n} (1-u_k) \geq 0;$$

whence the result follows. ■

Corollary 2.4 *Let $\{C_{ij} : 1 \leq i < j \leq n\}$ be a set of $\binom{n}{2}$ 2-copulas such that are PQD. Then the n-copula defined by* (2.1) *via Theorem* 2.1 *is POD.*

We now measure the dependence of any distribution constructed via Theorem 2.1. For that, we use a multivariate version of the well-known Spearman's rho coefficient which, for any n-copula C, is given by

$$\rho_n(C) = \frac{n+1}{2^n-(n+1)}\left[2^{n-1}\left(\int_{[0,1]^n} C(\mathbf{u})\mathrm{d}\Pi^n(\mathbf{u}) + \int_{[0,1]^n} \Pi^n(\mathbf{u})\mathrm{d}C(\mathbf{u})\right) - 1\right] \tag{2.4}$$

(Nelsen, 2002).

Theorem 2.5 *Let C be an n-copula given by* (2.1) *via Theorem* 2.1*. Then*

$$\rho_n(C) = \frac{n+1}{3[2^n-(n+1)]} \cdot \sum_{1\leq i<j\leq n} \rho(C_{ij}),$$

where $\rho(C_{ij})$ denotes the Spearman's rho coefficient associated with C_{ij}.

Proof. It is easy to check that for any n-copula D we have that

$$\int_{[0,1]^n} \Pi^n(\mathbf{u})\mathrm{d}D(\mathbf{u}) = \int_{[0,1]^n} \overline{D}(\mathbf{u})\mathrm{d}\Pi^n(\mathbf{u}).$$

Using Lemma 2.2, the expression (2.4), and some elementary algebra, we obtain the result. ■

3 Examples

Observe that if, at least, $\binom{n}{2} - 1$ bivariate margins in Theorem 2.1 are Π^2, then C is always an n-copula; and if all the bivariate margins are Π^2, then $C = \Pi^n$. In the following we provide examples using our method.

Example 3.1. Let C_{12}, C_{13} and C_{23} be the 2-copulas given by $C_{12}(u_1, u_2) = u_1u_2$, $C_{13}(u_1, u_3) = \alpha u_1u_3 + (1-\alpha)M(u_1, u_3)$ and $C_{23}(u_2, u_3) = \beta u_2u_3 + (1-\beta)W(u_2, u_3)$, (u_1, u_2, u_3) in $[0,1]^3$, where α and β are in $[0,1]$, and M and W denote the respective 2-copulas given by $M(u_1, u_3) = \min(u_1, u_3)$ and $W(u_2, u_3) = \max(u_2 + u_3 - 1, 0)$. Then the function C defined by $C(u_1, u_2, u_3) = (\alpha + \beta - 1)u_1u_2u_3 + (1-\alpha)u_2M(u_1, u_3) + (1-\beta)u_1W(u_2, u_3)$ via Theorem 2.1 is a 3-copula if and only if

$$(\alpha+\beta-1) + \frac{(1-\alpha)V_M([u_1, v_1] \times [u_3, v_3])}{(v_3 - u_3)(v_1 - u_1)} + \frac{(1-\beta)V_W([u_2, v_2] \times [u_3, v_3])}{(v_3 - u_3)(v_2 - u_2)} \geq 0.$$

Thus a sufficient condition for C to be a 3-copula is that $\alpha + \beta \geq 1$. Moreover, since C_{13} is PQD for all $\alpha \in [0,1]$ and C_{23} is PQD if and only if $\beta = 1$, we conclude that C is POD when $\beta = 1$. Finally, it is easy to check that $\rho_3(C) = (\beta - \alpha)/3$.

Example 3.2. Consider the well-known Farlie-Gumbel-Morgenstern (FGM) family of 2-copulas given by $C_\lambda(u, v) = uv[1 + \lambda(1-u)(1-v)]$, $(u, v) \in [0,1]^2$, with $\lambda \in [-1, 1]$. Let $\{C_{\lambda_{ij}} : 1 \leq i < j \leq n\}$ be a set of $\binom{n}{2}$ FGM 2-copulas such that λ_{ij} is in $[-1, 1]$. From (2.3), it is easy to check that the function C defined by (2.1) is an n-copula if and only if $\sum_{1 \leq i<j \leq n} \lambda_{ij}(1-2u_i)(1-2u_j) \geq -1$. Thus, a sufficient condition is that $|\sum_{1 \leq i<j \leq n} \lambda_{ij}| \leq 1$. Furthermore, if $\lambda_{ij} \geq 0$ for all (i, j), then C_{ij} is PQD and, as a consequence of Corollary 2.4, C is POD if $\sum_{1 \leq i<j \leq n} \lambda_{ij} \leq 1$. Finally, since $\rho(C_{ij}) = \lambda_{ij}/3$, then, from Theorem 2.5, we obtain that

$$\rho_n(C) = \frac{n+1}{9[2^n - (n+1)]} \cdot \sum_{1 \leq i<j \leq n} \lambda_{ij}.$$

As a particular case, if all the bivariate margins are $C_{ij}(u, v) = uv[1 + \lambda(1-u)(1-v)]$ for all (i, j), then the function C defined by (2.1) is an n-copula if and only if

$$-\frac{2}{n(n-1)} \leq \lambda \leq \frac{1}{\lfloor \frac{n}{2} \rfloor} \tag{3.1}$$

(see the technical lemma in Appendix), where $\lfloor x \rfloor$ denotes the integer part of the real number x; and, in such a case,

$$\rho_n(C) = \frac{n(n-1)(n+1)\lambda}{18[2^n - (n+1)]}.$$

We now compare these results with another known construction: An extension to a (2^n-n-1)-parameter FGM family of n-copulas. This extension is given by

$$E(\mathbf{u})=\prod_{i=1}^{n}u_i\left[1+\sum_{k=2}^{n}\sum_{1\le j_1<\cdots<j_k\le n}\theta_{j_1j_2\cdots j_k}\prod_{l=j_1}^{j_k}(1-u_l)\right]$$

(see Nelsen (1999) for details and references). After some calculus, we obtain that

$$\rho_n(E)=\frac{n+1}{2(2^n-n-1)}\cdot\sum_{k=2}^{n}\sum_{1\le j_1<\cdots<j_k\le n}\frac{1+(-1)^k}{3^k}\theta_{j_1j_2\cdots j_k}.$$

If, like our last case, all the bivariate margins—which belong to the FGM family of 2-copulas—are equal, i.e., they have the same parameter (θ), then

$$\rho_n(E)=\frac{(n+1)\theta}{2(2^n-n-1)}\cdot\sum_{k=2}^{n}\binom{n}{k}\frac{1+(-1)^k}{3^k}=\frac{(n+1)\theta}{2(2^n-n-1)}\left(\frac{4^n+2^n}{3^n}-2\right).$$

In the following, we show that the range of the Spearman's rho coefficient for the family of n-copulas of our example with a common parameter (λ) can be larger than the n-copulas given in the last paragraph. The minimum value for λ is $-2/[n(n-1)]$, while that the minimum value for θ is $-1/(2^n-n-1)$; whence

$$\rho_n(C)\ge-\frac{n+1}{9(2^n-n-1)}\quad\text{and}\quad\rho_n(E)\ge-\frac{(n+1)}{2(2^n-n-1)^2}\left(\frac{4^n+2^n}{3^n}-2\right)$$

for all $n\ge2$. If $n=2$, the bounds coincide, and, by induction, it is easy to check that

$$-\frac{n+1}{9(2^n-n-1)}<-\frac{(n+1)}{2(2^n-n-1)^2}\left(\frac{4^n+2^n}{3^n}-2\right)$$

for all $n\ge3$.

Appendix

Technical lemma. *Let* $\mathbf{a}=(a_1,a_2,\ldots,a_n)$ *be a point in* $\mathbb{R}^n$ *such that* $a_i\in\{-1,+1\}$ *for all* $i=1,2,\ldots,n$. *Then*

$$-\left\lfloor\frac{n}{2}\right\rfloor\le\sum_{1\le i<j\le n}a_ia_j\le\frac{n(n-1)}{2}.$$

Proof. It is clear that $\max\sum_{1\le i<j\le n}a_ia_j$ in the set $\{\mathbf{a}\in\mathbb{R}^n:\sum_{1\le i<j\le n}a_ia_j\ge0\}$ is obtained when $a_i=1$ for all $i=1,2,\ldots,n$, or $a_i=-1$ for all $i=1,2,\ldots,n$, and

is equal to $n(n-1)/2$. For the lower bound, we must find the minimum value of $\sum_{1\le i<j\le n} a_i a_j$ in the set $\{\mathbf{a} \in \mathbb{R}^n : \sum_{1\le i<j\le n} a_i a_j \le 0\}$. First, note that

$$\sum_{1\le i<j\le n} a_i a_j = \frac{1}{2}\sum_{i=1}^{n} a_i \sum_{\substack{j=1\\ j\neq i}}^{n} a_j .$$

Now, let k be the number of $a_i's$ such that are equal to 1, and let $n-k$ be the number of $a_i's$ such that are equal to -1 . Then, we have that $\sum_{i=1}^{n} a_i = 2k-n$. Thus,

$$\sum_{1\le i<j\le n} a_i a_j = \frac{1}{2}\left[(2k-n)\sum_{i=1}^{n} a_i - \sum_{i=1}^{n} a_i^2\right] = \frac{(2k-n)^2-n}{2},$$

and hence

$$\min \sum_{1\le i<j\le n} a_i a_j = \begin{cases} -\dfrac{n}{2}, & \text{if } n \text{ is even} \\ -\dfrac{n+1}{2}, & \text{if } n \text{ is odd} \end{cases} = -\left\lfloor \frac{n}{2} \right\rfloor,$$

which completes the proof. ■

Acknowledgements

We thank an anonymous referee for several comments. The first author thanks the research council of Yazd University for support. The second author also thanks the support by the Ministerio de Ciencia y Tecnología (Spain) and FEDER, under research project BFM2003-06522.

(Received February, 2005. Accepted June, 2005.)

References

Cohen, L. (1984). Probability distributions with given multivariate marginals. *J. Math. Phys.*, **25**, 2402–2403.

Cuadras, C. (1992). Probability distributions with given multivariate marginals and given dependence structure. *J. Multivariate Anal.*, **42**, 51–66.

Dall'Aglio, G. (1972). Fréchet classes and compatibility of distribution functions. *Symp. Math.*, **9**, 131–150.

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

Li, H., Scarsini, M. and Shaked, M. (1996). Linkages: a tool for the construction of multivariate distributions with given nonoverlapping multivariate marginals. *J. Multivariate Anal.*, **56**, 20–41.

Li, H., Scarsini, M. and Shaked, M. (1999). Dynamic linkages for multivariate distributions with given nonoverlapping multivariate marginals. *J. Multivariate Anal.*, **68**, 54–77.

Marco, J. M. and Ruiz-Rivas, C. (1992). On the construction of multivariate distributions with given nonoverlapping multivariate marginals. *Statist. Probab. Lett.*, **15**, 259–265.

Nelsen, R. B. (1999). *An Introduction to Copulas.* New York: Springer.

Nelsen, R. B. (2002). Concordance and copulas: A survey. In: *Distributions with Given Marginals and Statistical Modelling* (C. Cuadras, J. Fortiana and J. A. Rodríguez, eds.). Dordrecht: Kluwer Academic Publishers, pp. 169–178.

Rüschendorf, L. (1985). Construction of multivariate distributions with given marginals. *Ann. Inst. Statist. Math.*, **37**, 225–233.

Ali Dolati
Department of Statistics
Faculty of Science
Yazd University
Yazd, 89195-741, Iran
E-mail: dolati50@yahoo.com

Manuel Úbeda-Flores
Departamento de Estadística y Matemática Aplicada
Universidad de Almería
04120 Almería, Spain
E-mail: mubeda@ual.es