KYBERNETIKA — VOLUME 44 (2008), NUMBER 6, PAGES 846 – 863

SYMMETRIES OF RANDOM DISCRETE COPULAS

Arturo Erdely, José M. González–Barrios and Roger B. Nelsen

In this paper we analyze some properties of the discrete copulas in terms of permutations. We observe the connection between discrete copulas and the empirical copulas, and then we analyze a statistic that indicates when the discrete copula is symmetric and obtain its main statistical properties under independence. The results obtained are useful in designing a nonparametric test for symmetry of copulas.

Keywords: discrete copulas, r-symmetric permutations, independence

AMS Subject Classification: 60C05, 62E15, 62H05

1. INTRODUCTION

In Mayor et al. [8] and Mayor et al. [9] a class of binary aggregation operators on finite settings is studied, that is, discrete copulas. Let us start by defining a discrete copula on the finite chain $L = \{0, 1, \dots, n\}$.

Definition 1.1. A discrete copula C on L is a binary operation on L, i. e., $C : L \times L \to L$ satisfying the following properties:

i) $C(i, 0) = C(0, j) = 0$ for every $i, j \in L$.

ii) $C(i, n) = C(n, i) = i$ for every $i \in L$.

iii) If $0 \le i \le i' \le n$ and $0 \le j \le j' \le n$, then

$$C(i', j') - C(i', j) - C(i, j') + C(i, j) \ge 0,$$

that is, C is 2-increasing.

Here we observe that if we rescale the chain L to be $L' = \{0, 1/n, \dots, n/n = 1\}$, then Definition 1.1 agrees with the usual definition of subcopulas with domain $L' \times L' \subset [0, 1]^2$ when the range is also L', see for example Nelsen [12]. Therefore, a discrete copula C satisfies all the known properties of subcopulas. Definition 2 of discrete copulas in Kolesárová et al. [6], for the case $n = m$, coincides with rescaling the chain L to be $L' = \{0, 1/n. \dots, n/n = 1\}$, this definition is also used in Aguiló et al. [1] and Mesiar [10].

The use of discrete copulas can be related to observed data as noticed in Mesiar [10]. Recall that a binary operator C on the chain L is *symmetric or commutative* if and only if $C(i,j) = C(j,i)$ for every $i,j \in L$, and C is *associative* if and only if $C(C(i,j),k) = C(i,C(j,k))$ for every $i,j,k \in L$, see for example Alsina et al. [2], Klement et al. [4], Klement and Mesiar [5] or Schweizer and Sklar [13].

There exists a bijection between the set of $n \times n$ permutation matrices and the set of all discrete copulas on L, given in Proposition 6 and Corollary 1 of Mayor et al. [8], that states that C is a discrete copula if and only if there exists $A = (a_{ij})_{i,j\in\{1,\dots,n\}}$ a permutation matrix such that for every $r,s \in L$,

$$C(r,s) = \begin{cases} 0, & \text{if} \quad r = 0 \text{ or } s = 0 \\ \sum_{i\le r, j\le s} a_{ij} & \text{otherwise.} \end{cases}$$

From this result is easy to see that a discrete copula is symmetric or commutative if and only if its associated permutation matrix is symmetric, and that the number of discrete copulas on the chain L is $n!$. Also, if we define the $n \times n$ Łukasiewicz permutation matrix by $A = (a_{ij})_{i,j\in\{1,\dots,n\}}$, where $a_{ij} = 1$ if $i+j = n+1$, and $a_{ij} = 0$ otherwise, then a discrete copula C is associative if and only if C is an ordinal sum of Łukasiewicz matrices as proved in Proposition 9 in Mayor et al. [8]. Using the last result Kolesárová and Mordelová [7] observed using idempotent elements, that there are 2^{n-1} associative discrete copulas on L. It also follows that any associative discrete copula is necessarily symmetric or commutative. In Kolesárová et al. [6] it is proved that any discrete copula on $L' \times L'$ is a convex sum of irreducible discrete copulas.

In Section 2 of this paper we analyze in detail the connection between empirical copulas and discrete copulas via permutations, extending results of symmetries to r-symmetries. We make some observations about the order of the permutations and Landau's formula, we propose an associativity measure and find a nice geometric interpretation of associative samples.

In Section 3 we analyze a random sample measure of symmetry, studying some of its statistical properties under independence in terms of permutations.

In Section 4 we propose a symmetry test based on the statistic defined in Section 3. Many authors have tried to fit an Archimedean copula C to a data set. Recall that C is Archimedean if it is associative and its diagonal section $\delta_C(u) = C(u,u) < u$ for every $u \in (0,1)$. We also know that C must be symmetric, see Theorems 4.1.5 and 4.1.6 in Nelsen [12]. As far as we know there is no test for associativity of random data. However, if we test some random data for symmetry and we reject the hypothesis, then it does not seem appropriate to try to fit an Archimedean copula to these data, this provides a motivation for proposing a symmetry test. Finally, we include some final remarks.

2. EMPIRICAL COPULAS, DISCRETE COPULAS AND PERMUTATIONS

In this section we will think of a copula as a bivariate cumulative distribution function with uniform marginals, using Sklar's Theorem, see Nelsen [12].

We will begin this section by recalling the definition of the *empirical copula*, see for example Deheuvels [3]. Let us denote by $X_{[i]}$ and $Y_{[j]}$ the order statistics of a continuous random sample $(X_1, Y_1), \ldots, (X_n, Y_n)$ of a copula C, the empirical copula is defined by

$$C_n\left(\frac{i}{n}, \frac{j}{n}\right) = \frac{\text{num. of pairs } (X,Y) \text{ in the sample such that } X \le X_{[i]} \text{ and } Y \le Y_{[j]}}{n},$$

and we define $C_n(0, j/n) = 0 = C_n(i/n, 0)$, C_n clearly satisfies the 2-increasing condition, according to the definition of subcopula, see Nelsen [12]. Without losing generality we will always assume that $X_1 < X_2 < \cdots < X_n$, that is the order statistic $X_{[i]} = X_i$ for every $i = 1, 2, \ldots, n$. We also observe that for any $i, j \in \{1, 2, \ldots, n\}$, $C_n(i/n, j/n) = k/n$ for some $k = 0, 1, \ldots, n$.

In fact, since the empirical copula is invariant under strictly increasing transformations, we can assume that $X_1 = 1/n, X_2 = 2/n, \ldots, X_n = n/n = 1$, and that for every $k \in \{1, 2, \ldots, n\}$ there exists $j \in \{1, 2, \ldots, n\}$ such that $Y_k = j/n$. Even more, since the term $1/n$ is just a normalizing factor, we can assume that $X_1 = 1, X_2 = 2, \ldots, X_n = n$ and the values of Y are simply a permutation σ of $\{1, 2, \ldots, n\}$, that is $\sigma(i) = Y_i$ for $i = 1, 2, \ldots, n$. Therefore, from now on we will study a totally equivalent form of the empirical copula given by

$$C'_n(i,j) = \text{ num. of pairs } (X,Y) \text{ in the sample such that } X \le i \text{ and } Y \le Y_{[j]}, \quad (1)$$

where the sample is given by $(1, \sigma(1) = Y_1), (2, \sigma(2) = Y_2), \ldots, (n, \sigma(n) = Y_n)$, $(\sigma(1), \sigma(2), \ldots, \sigma(n))$ is a permutation σ of $\{1, 2, \ldots, n\}$ and $C'_n(i, 0) = 0 = C'_n(0, j)$. Besides, obviously $C'_n(i, n) = i$ and $C'_n(n, j) = j$. This approach will facilitate the study of several properties of the empirical copula in terms of permutations of $\{1, 2, \ldots, n\}$. In fact, with this definition the equivalent version of the empirical copula C'_n is simply a discrete copula on the chain L by Definition 1.1. Therefore, the representation of discrete copulas in terms of permutation matrices applies to C'_n, and it also gives a trivial proof of this characterization of discrete copulas. Of course all the properties stated in the introduction also follow for the empirical copula C'_n.

Now let us recall some basic notation for permutations. We know that a permutation σ on $I_n = \{1, 2, \ldots, n\}$ is the range of a bijection σ from I_n onto itself. Equivalently, a permutation is rearrangement of the elements of the ordered list I_n. In this paper we will denote a permutation σ of I_n as an ordered n-tuple of the form $(\sigma(1), \sigma(2), \ldots, \sigma(n))$. For example $(4, 2, 1, 3)$ is the permutation of I_4 such that $\sigma(1) = 4, \sigma(2) = 2, \sigma(3) = 1$ and $\sigma(4) = 1$. It is customary to use cycle notation in order to represent permutations. A *permutation cycle or orbit* is a subset of a permutation whose elements trade places with one another. For example in the permutation of I_4 given above, since σ is such that $1 \to 4 \to 3 \to 1$, and $2 \to 2$, then we have two cycles, which we will denote by $(1, 4, 3)$ and (2), the first one is a 3-cycle and the the second one a 1-cycle. Therefore we can write σ in terms of its cycles, in our example $\sigma = (1, 4, 3)(2)$. In many instances, the 1-cycles are omitted in these expressions. In general the number of elements in a cycle determine its order, so a k-cycle in a permutation σ on I_n is of the form $(i_1, i_2, \cdots, i_k)$, where $i_1, i_2, \ldots, i_k \in I_n$ and $1 \le k \le n$.

It is well known, see for example Skiena [14], that every permutation σ of I_n can be uniquely expressed as a product of disjoint cycles.

Using the concept of cycles, the theory of permutations has been immersed in group theory. In fact, since a permutation σ is a bijection on I_n, we can compose σ with itself to obtain powers of σ, where $\sigma^r = \sigma \circ \sigma^{r-1}$ for each $r \geq 2$.

Definition 2.1. The order of a permutation σ of I_n is the minimum value of $r \in \{1, 2, \ldots\}$ such that $\sigma^r = I$, where I is the identity function on I_n.

It is well known that any permutation σ of I_n has finite order and in fact, the order of a permutation is determined by the order of its cycles. The idea of the following result relies on the fact that a cycle of the form $(i_1, i_2, i_3, \cdots, i_r)$ can be written as $(i_1, \sigma(i_1), \sigma^2(i_1), \cdots, \sigma^{r-1}(i_1))$, where $\sigma^r(i_1) = i_1$.

Remark 2.2. Let σ be a permutation of I_n with unique disjoint cycle decomposition

$$\sigma = (i_{1,1}, i_{1,2}, \cdots, i_{1,r_1})(i_{2,1}, i_{2,2}, \cdots, i_{2,r_2}) \cdots (i_{k,1}, i_{k,2}, \cdots, i_{k,r_k}).$$

Then the order of σ is given by

$$\text{order}(\sigma) = \text{Least Common Multiple}\{r_1, r_2, \ldots r_k\}.$$

We will denote by LCM the least common multiple.

Now, let us recall that a sample $(X_1, Y_1), \ldots, (X_n, Y_n)$ is *symmetric* if and only if for every $i = 1, 2, \ldots, n$ if (X_i, Y_i) is in the sample then (Y_i, X_i) is also in the sample. Because we have agreed that our samples can be written as $(1, \sigma(1)), \ldots, (n, \sigma(n))$, then our samples are symmetric if and only if for each $i = 1, 2, \ldots, n$ if $(i, \sigma(i))$ is in the sample, so is $(\sigma(i), i)$. Therefore, for every $i = 1, 2, \ldots, n$, $\sigma^2(i) = i$. Hence, a sample is symmetric if and only if the permutation it generates is of order two, this fact was proved in Mayor et al. [8].

In some applications it is important to consider permutations σ of order r, for $r > 2$. We will say that a discrete copula C is r-symmetric if and only if its associated permutation σ has order r. In fact, we know that if C'_n is associative, then it is symmetric, so if σ is the permutation associated to C'_n, its order is 2. However, not every permutation σ of order 2 induces an associative discrete copula, see Example 2.5. It is easy to see how many r-symmetric discrete copulas exist, at least in the case when r is a prime number, maybe this result is known, but we do not have a reference, and we include its proof for completeness.

Proposition 2.3. The total number of r-symmetric samples of size n, $\sigma_r^{(n)}$, different from the identity, where r is a prime number is given by

$$\sigma_r^{(n)} = \sum_{k=0}^{[n/r]-1} \frac{n!}{(n-rk-r)!} \cdot \frac{1}{r^{k+1}(k+1)!}.$$

Proof. Let $(1, \sigma(1)), \ldots, (n, \sigma(n))$ be a r-symmetric sample, for some r prime integer. Then the order$(\sigma) = r$, according to Remark 2.2, if σ has l disjoint cycles, with cardinality $r_1, r_2, \ldots, r_l$, then $r = \mathrm{LCM}\{r_1, r_2, \ldots, r_l\}$. Therefore, there exists at least one $j \in \{1, 2, \ldots, l\}$ such that $r_j = r$, and the remaining $r_i's$ are equal to one or r. Therefore, $\sigma = (i_{1,1} \cdots i_{1,r})(i_{2,1} \cdots i_{2,r}) \cdots (i_{k,1} \cdots i_{k,r})$, where we have omitted all the cycles of order one, and $k \leq [n/r]$. Hence, if σ has k r-cycles, all we have to do is to select the r numbers that form each of those cycles, and observe that having selected r numbers, the number of different r-cycles is determined by $(r-1)!$, according to the order inside this cycle. Therefore,

$$\begin{aligned} \sigma_r^{(n)} &= \sum_{k=0}^{[n/r]-1} \binom{n}{r}\binom{n-r}{r}\cdots\binom{n-rk}{r}((r-1)!)^{k+1}/(k+1)! \\ &= \sum_{k=0}^{[n/r]-1} \frac{n(n-1)\cdots(n-rk)\cdots(n-rk-r+1)((k-1)!)^{k+1}}{(r!)^{k+1}(k+1)!} \\ &= \sum_{k=0}^{[n/r]-1} \frac{n!}{(n-rk-r)!}\cdot\frac{1}{r^{k+1}(k+1)!}, \end{aligned}$$

where the division by $(k+1)!$ is due to the fact that the order in which we select the k r-cycles is irrelevant. □

Observe that this formula applies to $r = 2$, extending the result in Mayor et al. [8]. In Table 1 we give the number of permutations of order k, $k = 2, 3, 5, 7$ for I_n with $n = 2, 3, \ldots, 15$, it can be seen that this numbers increases very rapidly for fixed k as n increases.

In the case in which k is not a prime number similar arguments can be used, but we have to be careful, since many other cases appear. For example if $k = 6$, we may have one or more 6-cycles, or we may have at least one 3-cycle and a 2-cycle, in order to have that the permutation has order $k = 6$.

Also, if k is not a prime number we can use a known recurrent formula to obtain $\sigma_2^{(k)}$, given by

$$\sigma_2^{(k)} = \sigma_2^{(k-1)} + (k-1)(\sigma_2^{(k-2)} + 1) \quad \text{with} \quad \sigma_2^{(2)} = 1, \sigma_2^{(3)} = 3.$$

Another interesting question is the following: What is the highest order of a permutation σ of I_n? In order to answer this question we have to refer to the concept of *integer partition of* n, that is the different ways in which n can be expressed as a sum of positive integers, for example the 5 integer partitions of 4 are

1.– 4
2.– $3+1$
3.– $2+2$
4.– $2+1+1$
5.– $1+1+1+1$

Table 1. Number of Permutations of Order k in I_n.

value of n	$\sigma_2^{(n)}$	$\sigma_3^{(n)}$	$\sigma_5^{(n)}$	$\sigma_7^{(n)}$
2	1	0	0	0
3	3	2	0	0
4	9	8	0	0
5	25	20	24	0
6	75	80	144	0
7	231	350	504	720
8	763	1232	1344	5760
9	2619	5768	3024	25920
10	9495	31040	78624	86400
11	35695	142010	809424	237600
12	140151	776600	4809024	570240
13	568503	4874012	20787624	1235520
14	2390479	27027728	72696624	892045440
15	10349539	168369110	1961583624	13348249200

Of course, we can identify the orders of the cycles of any permutation with a unique integer partition. Now, if σ is a permutation of I_n, and the integer partition generated by its r cycles is $k_1 + k_2 + \cdots + k_r$, where $k_1 \geq k_2 \geq \cdots \geq k_r$, and $k_1 + k_2 + \cdots + k_r = n$, then the order of σ equals $\mathrm{LCM}\{k_1, k_2, \cdots, k_r\}$. So, given any integer partition of n, $P(k_1, \ldots, k_r) = k_1 + k_2 + \cdots + k_r$, let

$$o(P(k_1, \ldots, k_r)) = \mathrm{LCM}\{k_1, k_2, \cdots, k_r\},$$

Then, we have the following bounds for the order of any permutation σ of I_n:

$$1 \leq \text{order of } \sigma \leq \max_{\{P_{k_1,\ldots,k_r} \text{integer partition of } n\}} o(P(k_1, \ldots, k_r)) = g(n). \qquad (2)$$

For example, for $n = 8$, the set

$$\{o(P(k_1, \ldots, k_r)) \mid P(k_1, \ldots, k_r) \text{ is an integer partition of } n\}$$

$$= \{1, 2, 3, 4, 5, 6, 7, 8, 10, 12, 15\}.$$

where 15 is attained at the integer partition $P(5, 3)$. Hence the maximum order of a permutation of I_8 is $g(8) = 15$.

The function $g(n)$ is known as Landau's formula. He proved that

$$\lim_{n\to\infty} \frac{\ln(g(n))}{\sqrt{n \ln(n)}} = 1.$$

Some interesting remarks about $g(n)$ can be found for example in Miller [11], as far as we know there is no closed formula for the values of $g(n)$.

Definition 2.4. Let $\underline{X_n} = (X_1, Y_1), (X_2, Y_2), \ldots, (X_n, Y_n)$ be a random sample of a copula C, with empirical copula C_n'. We define the *associativity measure* $A_{\underline{X_n}}$ of this sample by:

$$A_{\underline{X_n}} = \sum_{i,j,k \in \{1,2,\ldots,n-1\}} |C_n'(C_n'(i,j),k) - C_n'(i, C_n'(j,k))|.$$

First, we observe that $A_{\underline{X_n}}$ is well defined by equation (1) and the fact that the empirical copula C_n' take values of the form k, for $k = 0, 1, \ldots, n$. Besides, $A_{\underline{X_n}} = 0$ if and only if the sample is such that C_n' (equivalently C_n) is associative, in that case we will say that the *sample is associative*. Second, in the definition of $A_{\underline{X_n}}$ we do not include $i = n, j = n$ or $k = n$, because if $i = n, j = n$ or $k = n$ then $C_n'(C_n'(i,j),k) - C_n'(i, C_n'(j,k)) = 0$. There seems to be a direct relation between the order of a permutation associated to C_n' and the value of $A_{\underline{X_n}}$. Now, we will provide an example of a symmetric discrete copula which is non associative.

Example 2.5. Let $\underline{X_n} = \{(X_1, Y_1), \ldots (X_n, Y_n)\}$ be a sample of size n from a copula C. Assume that the first coordinate is such that $0 < X_1 < X_2 < \cdots < X_n$ and $Y_1 = Y_{[n]}$, $Y_j = Y_{[j]}$, for $j = 2, 3, \ldots, n-1$ and $Y_n = Y_{[1]}$. Then $\underline{X_n}$ is symmetric and

$$A_{\underline{X_n}} = 2 \sum_{k=3}^{n-1} \binom{k-1}{2} + 2 \sum_{j=4}^{n-1} \binom{j-2}{2}.$$

Since the sample is given by

$$\underline{X_n} = \{(X_{[1]}, Y_{[n]}), (X_{[2]}, Y_{[2]}), \ldots, (X_{[n-1]}, Y_{[n-1]}), (X_{[n]}, Y_{[1]})\}$$

it is clearly symmetric, see Figure 1.

Equivalently, using equation (1), the sample can be thought as

$$\underline{X_n} = \{(1, n), (2, 2), \ldots, (n-1, n-1), (n, 1)\}.$$

From here, we have that the discrete copula $C_n^{'}$ equivalent to the empirical copula is given by

$$C_n^{'}(i,j) = \begin{cases} \min\{i,j\} - 1 & \text{if } \ 1 \le i, j < n \\ \min\{i,j\} & \text{if } \ i = n \text{ or } j = n. \end{cases}$$

Observe that if i or j are equal to 1, then $C_n^{'}(i,j) = 0$, except for the case in which the other one is n. Now, since

$$A_{\underline{X_n}} = \sum_{i,j,k \in \{1,2,\ldots,n-1\}} |C_n'(C_n'(i,j),k) - C_n'(i, C_n'(j,k))|,$$

it is easy to evaluate $A_{\underline{X_n}}$ proceeding by cases.

Since associative discrete copulas are ordinal sums of Łukasiewicz permutation matrices, as noticed in Mayor et al. [8] and Kolesárová and Mordelová [7]. We do have a very simple and nice geometric representation of associativity in terms of an

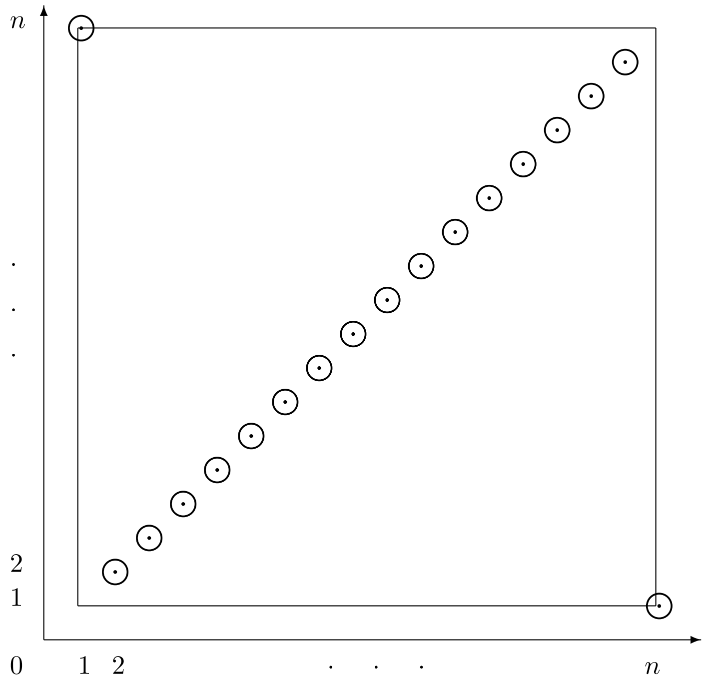

Fig. 1. Symmetric sample $\underline{X_n}$ with large value of $A_{\underline{X_n}}$.

associative sample and its idempotent elements. See Figure 2, where an associative sample of size $n = 18$ is shown in the case that the idempotent elements are $i_0 = 0, i_1 = 3, i_2 = 7, i_3 = 8, i_4 = 9, i_5 = 10, i_6 = 11, i_7 = 12$ and $i_8 = 18 = n$. That is from rescaling the original sample into $(1, \sigma(1)), \dots (n, \sigma(n))$ and observing the resulting graph we can easily deduce if C'_n is associative or not, just by checking if all points are located in the main diagonal or in secondary diagonals.

3. A NEW SYMMETRY STATISTIC OF A COPULA

We have already seen that a sample of the form $(1, \sigma(1)), (2, \sigma(2)), \dots, (n, \sigma(n))$, where σ is a permutation on I_n, is symmetric if and only if $\sigma^2(i) = i$ for every $i \in I_n$. It is natural then to define a sample measure of symmetry by taking:

$$S_\sigma^n = \sum_{i=1}^{n} \left(i - \sigma^2(i)\right)^2 . \tag{3}$$

Of course $S_\sigma^n = 0$ if and only if the sample is symmetric. In fact, under the hypothesis of independence, all $n!$ permutations σ of I_n are equally probable. Therefore, we can find the exact probability of $\sigma(i) = j$ for any $i, j \in I_n$.

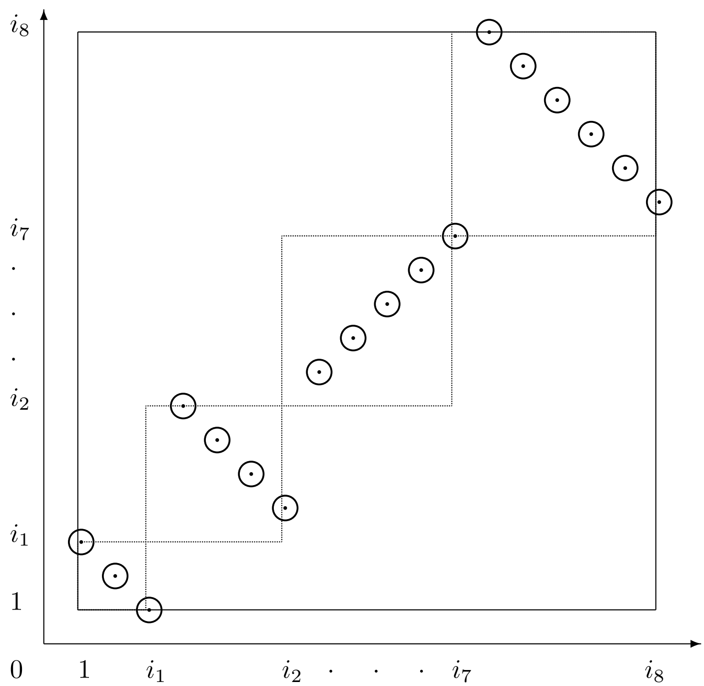

Fig. 2. Associative sample with idempotent elements $i_0 = 0 < i_1 < \cdots < i_8 = n = 18$.

Theorem 3.1. Let $\underline{X_n} = (X_1, Y_1), (X_2, Y_2), \ldots, (X_n, Y_n)$ be a continuous random sample of size n where X and Y are independent and continuous random variables, that is, the pair (X, Y) has copula Π. Let $(1, \sigma(1)), (2, \sigma(2)), \ldots, (n, \sigma(n))$ be the modified sample that generates Π'_n. Then for any $i \in I_n$ and σ a random permutation of I_n we have that

$$\Pr\left(\sigma^2(i) = i\right) = \frac{2}{n}$$

and

$$\Pr\left(\sigma^2(i) = j\right) = \frac{(n-2)}{n(n-1)} \quad \text{for any} \quad j \in I_n \backslash \{i\}.$$

Proof. If $(1, \sigma(1)), (2, \sigma(2)), \ldots, (n, \sigma(n))$ is the modified sample from the independent copula, the event $\{\sigma^2(i) = i\}$ can be written as

$$\{\sigma^2(i) = i\} = \{\sigma(i) = i\} \cup \left(\cup_{k=1, k \neq i}^{n} \{\sigma(i) = k, \sigma(k) = i\}\right),$$

that is, i belongs to a one-cycle of σ or i belongs to a two-cycle of σ. Now,

$$\Pr(\sigma(i) = i) = \frac{(n-1)!}{n!} = \frac{1}{n},$$

since fixing $\sigma(i) = i$, we can permute the remaining $n - 1$ elements of I_n in $(n-1)!$

ways. Also,

$$\Pr\left(\cup_{k=1,k\neq i}^{n}\{\sigma(i)=k,\sigma(k)=i\}\right)=\frac{\binom{n-1}{1}(n-2)!}{n!}=\frac{1}{n},$$

since we first select the value of k that forms the two-cycle together with i, and then the remaining $n-2$ values of I_n can be permuted in $(n-2)!$ ways. Hence

$$\Pr\left(\sigma^2(i)=i\right)=\frac{1}{n}+\frac{1}{n}=\frac{2}{n}.$$

Now let $i \in I_n$ and $j \in I_n$ different from i. Then the event $\{\sigma^2(i)=j\}$ can be written as

$$\{\sigma^2(i)=j\}=\cup_{k=1,k\neq i,j}^{n}\{\sigma(i)=k,\sigma(k)=j\},$$

because i and j must belong to at least a three-cycle of σ. Then

$$\Pr\left(\sigma^2(i)=j\right)=\frac{\binom{n-2}{1}(n-2)!}{n!}=\frac{n-2}{n(n-1)},$$

since first we have to select k from $I_n\backslash\{i,j\}$, then we have to fix $\sigma(i)$ as k and $\sigma(k)$ as j, the remaining values of σ can be permuted in $(n-2)!$ ways, including the value of $\sigma(j)$. □

Using Theorem 3.1 we can find the expectation of S_σ^n.

Proposition 3.2. For any $n \geq 2$ and under the hypothesis of independence.

$$\mathrm{E}\left(\left(\sigma^2(i)\right)^k\right)=\frac{n-2}{n(n-1)}\sum_{j=1}^{n} j^k+\frac{i^k}{n-1}.$$

Besides,

$$\mathrm{E}\left(S_\sigma^n\right)=\frac{(n-2)n(n+1)}{6}. \tag{4}$$

Proof. From Theorem 3.1

$$\begin{aligned}\mathrm{E}\left(\left(\sigma^2(i)\right)^k\right) &= \sum_{j=1}^{i-1} j^k\frac{n-2}{n(n-1)}+i^k\frac{2}{n}+\sum_{j=i+1}^{n} j^k\frac{n-2}{n(n-1)}\\ &= \sum_{j=1}^{n} j^k\frac{n-2}{n(n-1)}+i^k\frac{2}{n}-i^k\frac{n-2}{n(n-1)}\\ &= \frac{n-2}{n(n-1)}\sum_{j=1}^{n} j^k+\frac{i^k}{n-1}.\end{aligned}$$

Now, using the above expression for $k=1$ equation (10) and the fact that

$$\sum_{i=1}^{n}(\sigma^2(i))^2 = \sum_{i=1}^{n} i^2 = \frac{n(n+1)(2n+1)}{6},$$

we get

$$\begin{aligned}
\mathrm{E}\left(S_\sigma^n\right) &= E\left(\sum_{i=1}^{n} i^2 - 2\sum_{i=1}^{n} i\sigma^2(i) + \sum_{i=1}^{n}\left(\sigma^2(i)\right)^2\right) \\
&= \frac{n(n+1)(2n+1)}{3} - 2\sum_{i=1}^{n} i\left(\frac{n-2}{n(n-1)}\sum_{j=1}^{n} j + \frac{i}{n-1}\right) \\
&= \frac{n(n+1)(2n+1)}{3} - \frac{n(n+1)^2(n-2)}{2(n-1)} - \frac{2n(n+1)(2n+1)}{6(n-1)} \\
&= \frac{n(n+1)}{6(n-1)}\left(n^2-3n+2\right) \\
&= \frac{(n-2)n(n+1)}{6}.
\end{aligned}$$

□

In order to find the variance, we need to find first the joint distribution of $\sigma^2(i)$ and $\sigma^2(j)$ for $i \neq j$.

Theorem 3.3. Let $\underline{X_n} = (X_1,Y_1),(X_2,Y_2),\ldots,(X_n,Y_n)$ be a random sample of the product copula Π for $n \geq 5$. Let $(1,\sigma(1)),(2,\sigma(2)),\ldots,(n,\sigma(n))$ be the modified sample that generates C_n'. Then for any $i,j \in I_n$ with $i \neq j$ and σ permutation of I_n we have that

$$\Pr\left(\sigma^2(i)=k,\sigma^2(j)=l\right) = \begin{cases} 0 & \text{if } \; k=l\in\{1,2,\ldots,n\} \\ \frac{5}{n(n-1)} & \text{if } \; k=i \text{ and } l=j \\ \frac{1}{n(n-1)} & \text{if } \; k=j \text{ and } l=i \\ \frac{(2n-7)}{n(n-1)(n-2)} & \text{if } \; k=i \text{ and } l\neq i,j \\ \frac{(2n-7)}{n(n-1)(n-2)} & \text{if } \; k\neq i,j \text{ and } l=j \\ \frac{(n-3)}{n(n-1)(n-2)} & \text{if } \; k=j \text{ and } l\neq i,j \\ \frac{(n-3)}{n(n-1)(n-2)} & \text{if } \; k\neq i,j \text{ and } l=i \\ \frac{2(n-3)+(n-4)(n-5)}{n(n-1)(n-2)(n-3)} & \text{if } \; k\neq l \text{ and } k,l\neq i,j. \end{cases}$$

Proof. Let us denote the joint density of $\sigma^2(i)$ and $\sigma^2(j)$ for $i \neq j$, $i,j \in I_n$ by

$$f(k,l) = \Pr\left(\sigma^2(i)=k,\sigma^2(j)=l\right).$$

First we observe that since σ is a bijection on I_n, then $f(k,k)=0$ for every $k \in I_n$.

Second, to evaluate $f(i,j)$, we observe that we have five possibilities for the permutation σ. If (i) and (j) are one-cycles of σ, if (ij) is a two-cycle of σ, if (i) and (jr) are a one-cycle and a two-cycle of σ, if (j) and (ir) are a one-cycle and a two-cycle of σ, or if (ir) and (js) are two-cycles of σ. Then

$$\begin{aligned} f(i,j) &= \Pr(\sigma(i)=i, \sigma(j)=j) + \Pr(\sigma(i)=j, \sigma(j)=i) \\ &\quad + \Pr(\sigma(i)=i, \sigma(j)=r, \sigma(r)=j, r \neq i,j) \\ &\quad + \Pr(\sigma(j)=j, \sigma(i)=r, \sigma(r)=i, r \neq i,j) \\ &\quad + \Pr(\sigma(i)=r, \sigma(r)=i, \sigma(j)=s, \sigma(s)=j,\ r \neq s,\ r,s \neq i,j) \\ &= \frac{2(n-2)!}{n!} + \frac{2\binom{n-2}{1}(n-3)!}{n!} + \frac{2\binom{n-2}{2}(n-4)!}{n!} \\ &= \frac{5}{n(n-1)}. \end{aligned}$$

Third, to find $f(j,i)$, observe that the only possibility is that σ contains a four-cycle of the form $(irjs)$. Hence

$$\begin{aligned} f(j,i) &= \Pr(\sigma(i)=r, \sigma(r)=j, \sigma(j)=s, \sigma(s)=i) \\ &= \frac{2\binom{n-2}{2}(n-4)!}{n!} \\ &= \frac{1}{n(n-1)}. \end{aligned}$$

Fourth, in order to find $f(i,l)$ for $l \neq i,j$, we observe that we have two possibilities: If (i) and $(jsl\cdots)$ are a one-cycle and at least a three-cycle of σ, or if (ir) and $(jsl\cdots)$ are a two-cycle and at least a three-cycle of σ. Therefore,

$$\begin{aligned} f(i,l) &= \Pr(\sigma(i)=i, \sigma(j)=s, \sigma(s)=l,\ s \neq i,j,l) \\ &\quad + \Pr(\sigma(i)=r, \sigma(r)=i, \sigma(j)=s, \sigma(s)=l,\ r \neq s,\ r,s \neq i,j) \\ &= \frac{\binom{n-3}{1}(n-3)!}{n!} + \frac{2\binom{n-3}{2}(n-4)!}{n!} \\ &= \frac{2n-7}{n(n-1)(n-2)}. \end{aligned}$$

Fifth, to evaluate $f(k,j)$ for $k \neq i,j$, we simply observe that interchanging i and j we have the same result as in case four. So,

$$f(k,j) = \frac{2n-7}{n(n-1)(n-2)}.$$

Sixth, in order to find $f(j,l)$ for $l \neq i,j$, we observe that there are two possibilities:

If (ilj) is a three-cycle of σ, or if $(irjsl\cdots)$ is at least a five-cycle of σ. Then

$$\begin{aligned} f(j,l) &= \Pr(\sigma(i)=l,\sigma(l)=j,\sigma(j)=i) \\ &\quad + \Pr(\sigma(i)=r,\sigma(r)=j,\sigma(j)=s,\sigma(s)=l,\ r\neq s, r,s\neq i,j,l) \\ &= \frac{(n-3)!}{n!} + \frac{2\binom{n-3}{2}(n-4)!}{n!} \\ &= \frac{(n-3)}{n(n-1)(n-2)}. \end{aligned}$$

Seventh, to evaluate $f(k,i)$ for $k\neq i,j$, we simply observe that interchanging i and j we have the same result as in case six. So,

$$f(k,i)=\frac{(n-3)}{n(n-1)(n-2)}.$$

Eighth, in order to find $f(k,l)$ for $k\neq l$ and $k,l\neq i,j$, we observe that we have three possibilities: if $(ijkl\cdots)$ is at least a four-cycle of σ, if $(jilk\cdots)$ is at least a four-cycle of σ, or if $(irk\cdots)$ and $(jsl\cdots)$ are at least three-cycles of σ, observe that both can belong to the same cycle, that is it is possible that $(irk\cdots jsl)$ is a cycle of σ. Hence,

$$\begin{aligned} f(k,l) &= \Pr(\sigma(i)=j,\sigma(j)=k,\sigma(k)=l) \\ &\quad + \Pr(\sigma(i)=l,\sigma(l)=k,\sigma(j)=i) \\ &\quad + \Pr(\sigma(i)=r,\sigma(r)=k,\sigma(j)=s,\sigma(s)=l,\ r\neq s,\ r,s\neq i,j) \\ &= \frac{2(n-3)!}{n!} + \frac{2\binom{n-4}{2}(n-4)!}{n!} \\ &= \frac{2(n-3)+(n-4)(n-5)}{n(n-1)(n-2)(n-3)}. \end{aligned}$$

By summing over all pairs $f(k,l)$ we obtain

$$\begin{aligned} \sum_{k,l=1}^{n} f(k,l) &= \frac{(n-2)!}{n!}(5+1+2(2n-7)+2(n-3)+2(n-3)+(n-4)(n-5)) \\ &= \frac{(n-2)!}{n!}(n^2-n) \\ &= 1. \end{aligned}$$

Therefore, $f(k,l)$ is a density. $\square$

Now we are ready to compute the variance of S_σ^n.

Proposition 3.4. For any $n\geq 5$ and under the hypothesis of independence.

$$\mathrm{Var}(S_\sigma^n)=\frac{n(n+1)(5n^3+9n^2-12)}{180}.$$

If $n = 2$ then $\mathrm{Var}(S^2_\sigma) = 0$, and if $n = 3$ then $\mathrm{Var}(S^3_\sigma) = 8$.

P r o o f. From Proposition 3.2 we know that $\mathrm{E}(S^n_\sigma) = (n(n-2)(n+1)/6$. Hence

$$\mathrm{Var}(S^n_\sigma) \;=\; \mathrm{E}((S^n_\sigma)^2) - \left[\frac{n(n-2)(n+1)}{6}\right]^2.$$

Since $\quad S^n_\sigma = \sum_{i=1}^n [\, i - \sigma^2(i)\,]^2 = \dfrac{n(n+1)(2n+1)}{3} - 2\sum_{i=1}^n i\,\sigma^2(i) \quad$ then

$$\begin{aligned}(S^n_\sigma)^2 \quad &= \quad \left[\frac{n(n+1)(2n+1)}{3}\right]^2 - \frac{4}{3}\,n(n+1)(2n+1)\sum_{i=1}^n i\,\sigma^2(i) \\ &\quad + 4\Big[\,\sum_{i=1}^n i^2\sigma^2(i)^2 + \sum\sum_{i\neq j} ij\,\sigma^2(i)\,\sigma^2(j)\Big].\end{aligned}$$

Using the above expression for $(S^n_\sigma)^2$ we evaluate its expectation using Proposition 3.2 and Theorem 3.3 to obtain:

$$\mathrm{E}((S^n_\sigma)^2) = \left[\frac{n(n+1)(2n+1)}{3}\right]^2 - \frac{4}{3}\,n(n+1)(2n+1)\varphi_0(n) \;+ 4[\,\varphi_1(n) + \varphi_2(n)\,],$$

where

$$\varphi_0(n) := \sum_{i=1}^n i\,\mathrm{E}[\,\sigma^2(i)\,] = \frac{n(n-2)(n+1)^2}{4(n-1)} + \frac{n(n+1)(2n+1)}{6(n-1)}\,,$$

$$\begin{aligned}\varphi_1(n) \quad &:= \quad \sum_{i=1}^n i^2\,\mathrm{E}[\,\sigma^2(i)^2\,] \\ &= \quad \frac{n(n-2)(n+1)^2(2n+1)^2}{36(n-1)} + \frac{n(n+1)(2n+1)(3n^2+3n-1)}{30(n-1)},\end{aligned}$$

and

$$\begin{aligned}\varphi_2(n) \quad &:= \quad \sum\sum_{i\neq j} ij\,\mathrm{E}\,[\,\sigma^2(i)\,\sigma^2(j)\,]\Big] = \frac{6}{n(n-1)}\,g_1(n) + \frac{2(3n-10)}{n(n-1)(n-2)}\,g_2(n) \\ &\qquad + \frac{2(n-3)+(n-4)(n-5)}{n(n-1)(n-2)(n-3)}\,g_3(n)\end{aligned}$$

where after some tedious algebra we obtain

$$g_1(n) := \sum\sum_{i\neq j} i^2 j^2 = \Big(\sum i^2\Big)^2 - \sum i^4\,,$$

$$\begin{aligned}g_2(n) \quad &:= \quad \sum\sum\sum_{i\,,\,j\,,\,k \text{ different}} i^2 jk \\ &= \quad \Big(\sum i\Big)^2\Big(\sum i^2\Big) - \Big(\sum i^2\Big)^2 - 2\Big(\sum i\Big)\Big(\sum i^3\Big) + 2\sum i^4,\end{aligned}$$

and

$$g_3(n)=\sum\sum\sum\sum_{i\,,\,j\,,\,k\,,\,l\text{ different}} ijk\,l$$
$$=\Big(\sum i\Big)^4-6\sum i^4+3\Big(\sum i^2\Big)^2+8\Big(\sum i\Big)\Big(\sum i^3\Big)-6\Big(\sum i\Big)^2\Big(\sum i^2\Big)$$

using the well known formulas $\sum i = \frac{1}{2}\,n(n+1)$, $\sum i^2 = \frac{1}{6}\,n(n+1)(2n+1)$, $\sum i^3 = \frac{1}{4}\,n^2(n+1)^2$, $\sum i^4 = \frac{1}{30}\,n(n+1)(2n+1)(3n^2+3n-1)$. Making the substitutions and simplifying we obtain

$$\mathrm{Var}(S_\sigma^n)=\frac{n(n+1)(5n^3+9n^2-12)}{180}.$$

If $n=2$ then σ^2 is the identity, hence $\mathrm{Var}(S_\sigma^2)=0$, and for $n=3$ the density of S_σ^3 is given by

$$\Pr(S_\sigma^3=k)=\begin{cases} 2/3 & \text{if} \quad k=0 \\ 1/3 & \text{if} \quad k=6 \end{cases}$$

Therefore $\mathrm{Var}(S_\sigma^3)=8$. For $n=4$ if we let $f(k)=\Pr(S_\sigma^4=k)$, it easy to see that $f(0)=5/12$, $f(4)=f(16)=f(20)=1/12$ y $f(6)=f(14)=2/12$, hence $E(S_\sigma^4)=20/3$ and $\mathrm{Var}(S_\sigma^4)=452/9$, which finishes the proof. □

In Table 2, values for $\mathrm{E}(S_\sigma^n)$, $\sqrt{\mathrm{Var}(S_\sigma^n)}$, and $\sqrt{\mathrm{Var}(S_\sigma^n)}/\mathrm{E}(S_\sigma^n)$ are given for different values of n.

Since we have explicit expressions for the expectation and variance of S_σ^n (under independence), we can define S_σ^n in its standardized form

$$T_\sigma^n := \frac{\sum_{i=1}^n [\,i-\sigma^2(i)\,]^2 - n(n-2)(n+1)/6}{\sqrt{n(n+1)(5n^3+9n^2-12)/180}}\,, \tag{5}$$

so that $\mathrm{E}(T_\sigma^n)=0$ and $\mathrm{Var}(T_\sigma^n)=1$. We also observe that S_σ^n and T_σ^n are sums of identically distributed random variables, but the terms of this sum are not independent, see Theorems 3.1 and 3.3. Therefore its exact distribution is not easy to find. However, the statistics S_σ^n or T_σ^n can be used to test for symmetry as we will see in the next section.

4. A NEW SYMMETRY TEST FOR COPULAS AND FINAL REMARKS

In order to propose a test of symmetry, we could use the statistic S_σ^n or equivalently the standardized version T_σ^n, since they measure the symmetry of samples. Of course large values of these statistics give evidence of asymmetric densities.

So let us assume that $\underline{X_n}=(X_1,Y_1),\dots,(X_n,Y_n)$ is a random sample of a common unknown copula C, and let $(1,\sigma(1)),\dots,(n,\sigma(n))$ the rescaled sample defined in Section 2. In order to test the hypotheses

$$H_0 : C \text{ is a symmetric copula} \quad \text{vs} \quad H_1 : C \text{ is not a symmetric copula}, \tag{6}$$

at level $0<\alpha<1$. We propose the following methodology:

Table 2. Values for $\mathrm{E}(S_\sigma^n)$, $\sqrt{\mathrm{Var}(S_\sigma^n)}$, and $\sqrt{\mathrm{Var}(S_\sigma^n)}/\mathrm{E}(S_\sigma^n)$.

value of n	$\mathrm{E}(S_\sigma^n)$	$\left[\mathrm{Var}(S_\sigma^n)\right]^{1/2}$	$\left[\mathrm{Var}(S_\sigma^n)\right]^{1/2}/\mathrm{E}(S_\sigma^n)$
2	0.000	0.000	—
3	2.000	2.828	1.414
4	6.667	7.087	1.063
5	15.000	11.818	0.788
6	28.000	18.022	0.644
7	46.667	25.827	0.553
8	72.000	35.350	0.491
9	105.000	46.701	0.445
10	146.667	59.985	0.409
30	4340.000	859.819	0.198
50	20400.000	3028.654	0.148

- Obtain the value of S_σ^n (or T_σ^n) for the original sample.
- Simulate a large number m of samples of size n coming from the independent copula Π.
- Obtain for each simulated sample the value of S_σ^n (or T_σ^n).
- Estimate the quantile $(1-\alpha)$ of S_σ^n (or T_σ^n) using the m simulations.
- Reject H_0 if the value of S_σ^n (or T_σ^n) for the original sample exceeds the estimated quantile.

In order to clarify this proposal, we observe that we are simulating m samples coming from the independent copula Π, instead of sampling from the true copula C, but since C is unknown we can not simulate samples from it. The idea of simulating samples from the product copula, and comparing the $(1-\alpha)$-quantile of S_σ^n (or T_σ^n) with the respective statistic of the original sample, is based on the idea that the product copula may produce the greatest asymmetries among the family of symmetric copulas, this statement is not easy to prove. However, we observed this behavior when we simulated many sets of data from several known symmetric families. Hence, in case H_0 holds, our proposal makes comparisons against the "most extreme case" of symmetric copulas.

The methodology proposed above works nicely if the true copula is quite asymmetric, for example if the copula has support $S = ([1/3, 2/3] \times [0, 1/3]) \cup ([2/3, 1] \times [1/3, 2/3]) \cup ([0, 1/3] \times [2/3, 1])$ and it is uniform in each little square of its domain. However, if the copula is slightly asymmetric, it is difficult to detect this asymmetry using this methodology.

It is also important to notice that in the definition of the empirical copula C_n, its domain is the grid $\{0, 1/n, \ldots, (n-1)/n, 1\}^2$, and not the grid generated by the original sample $\underline{X_n} = (X_1, Y_1), \ldots, (X_n, Y_n)$. Hence, we can not recover the original sample if we only know its empirical copula.

That is symmetry of the empirical copula or the discrete copula does not necessarily imply symmetry of the sample.

Future work: It seems only natural to provide a method to test for associativity of a continuous random sample $(X_1, Y_1), \ldots, (X_n, Y_n)$.

ACKNOWLEDGEMENT

We want to thank the referees for a careful revision and interesting comments on the first version of the manuscript. This work was partially supported by Mexico's CONACyT grant 50152.

(Received May 13, 2008.)

REFERENCES

[1] I. Aguiló, J. Suñer, and J. Torrens: Matrix representation of discrete quasi-copulas. Fuzzy Sets and Systems *159* (2008), 1658–1672.

[2] C. Alsina, M. J Frank, and B. Schweizer: Associative Functions: Triangular Norms and Copulas. World Scientific Publishing Co., Singapore 2006.

[3] P. Deheuvels: La fonction de dépendance empirique et ses propriétés. Un test non paramétrique d'indépendance. Acad. Roy. Belg. Bull. Cl. Sci. *65* (1979), 5, 274–292.

[4] E. P. Klement, R. Mesiar, and E. Pap: Triangular Norms. Kluwer Academic Publishers, Dordrecht 2000.

[5] E. P. Klement and R. Mesiar: Logical, Algebraic, Analytic, and Probabilistic Aspects of Triangular Norms. Elsevier, Amsterdam 2005.

[6] A. Kolesárová, R. Mesiar, J. Mordelová, and C. Sempi: Discrete copulas. IEEE Trans. Fuzzy Systems. *14* (2006), 698–705.

[7] A. Kolesárová and J. Mordelová: Quasi-copulas and copulas on a discrete scale. Soft Computing *10* (2006), 495–501.

[8] G. Mayor, J. Suñer, and J. Torrens: Copula-like operations on finite settings. IEEE Trans. Fuzzy Systems *13* (2005), 468–477.

[9] G. Mayor, J. Suñer, and J. Torrens: Sklar's Theorem in finite settings. IEEE Trans. Fuzzy Systems *15* (2007), 410–416.

[10] R. Mesiar: Discrete copulas – what they are. In: Joint EUSFLAT-LFA 2005, Conference Proceedings (E. Montseny and P. Sobrevilla, eds.) Universitat Politecnica de Catalunya, Barcelona 2005, pp. 927–930.

[11] W. Miller: The maximum order of an element of a finite symmetric group. Amer. Math. Monthly *94* (1987), 6, 497–506.

[12] R. B. Nelsen: An Introduction to Copulas. Second edition. Springer, New York 2006.

[13] B. Schweizer and A. Sklar: Probabilistic Metric Spaces. North-Holland, New York 1983.

[14] S. Skiena: The cycle structure of permutations. In: Implementing Discrete Mathematics: Combinatorial and Graph Theory with Mathematica. Addison-Wesley, Reading, MA 1990, pp. 20–24.

Arturo Erdely, Universidad Anáhuac México Norte, Escuela de Actuaría, Av. Lomas Anáhuac s/n, C.P. 52786, Huixquilucan, Estado de México. México.

José M. González-Barrios: Universidad Nacional Autónoma de México, Instituto de Investigaciones en Matemáticas Aplicadas y en Sistemas, Departamento de Probabilidad y Estadística, Circuito Escolar s/n, Ciudad Universitaria, C.P. 04510, México D.F. México.

Roger B. Nelsen: Lewis & Clark College, Mathematical Sciences Department, 0615 SW Palatine Hill Rd., Portland OR, 97219. U. S. A.

e-mails: aerdely@anahuac.mx, gonzaba@sigma.iimas.unam.mx, nelsen@lclark.edu