Kybernetika

José M. González-Barrios

Statistical aspects of associativity for copulas

Kybernetika, Vol. 46 (2010), No. 1, 149--177

Persistent URL: http://dml.cz/dmlcz/140056

KYBERNETIKA — VOLUME *46* (2010), NUMBER 1, PAGES 149–177

STATISTICAL ASPECTS OF ASSOCIATIVITY FOR COPULAS

José M. González–Barrios

In this paper we study in detail the associativity property of the discrete copulas. We observe the connection between discrete copulas and the empirical copulas, and then we propose a statistic that indicates when an empirical copula is associative and obtain its main statistical properties under independence. We also obtained asymptotic results of the proposed statistic. Finally, we study the associativity statistic under different copulas and we include some final remarks about associativity of samples.

Keywords: discrete copulas, associativity, permutations, independence

Classification: 60C05, 62E15, 62H05

1. INTRODUCTION

The concept of discrete copulas defined in Mayor et al. [14] and Mayor et al. [15] as a class of binary aggregation operators on finite settings has been proved to be quite useful to study empirical copulas, see for example Erdely et al. [4]. 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 $0 \leq i \leq i' \leq n$ and $0 \leq j \leq j' \leq n$, then

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

that is, C is 2-increasing.

As observed in Erdely et al. [4], if we rescale the chain L to be $L' = \{0, 1/n, \dots \dots, n/n = 1\}$, then Definition 1.1 agrees with the usual definition of subcopulas with domain $L' \times L' \subset [0, 1]^2$, when the range is L', see for example Nelsen [17]. Definition 2 of discrete copulas in Kolesárová et al. [11], for the case $n = m$, coincides with rescaling the chain L to be $L' = \{0, 1/n. \ldots, n/n = 1\}$, this definition is also used in Aguiló et al. [1] and Mesiar [16].

As noticed in Mesiar [16], we can use discrete copulas to describe observed data. 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. [8], Klement and Mesiar [9] or Schweizer and Sklar [19].

Mayor et al. [14] proved the existence of a bijection between the set of $n \times n$ permutation matrices and the set of all discrete copulas on L, given in their Proposition 6 and Corollary 1, that states that C is a discrete copula if and only if there exists $A = (a_{ij})_{i,j\in\{1,\ldots,n\}}$ 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 $A = (a_{ij})_{i,j\in\{1,\ldots,n\}}$, 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. [14]. This observation is the main key of the present article. It also follows that any associative discrete copula is necessarily symmetric or commutative. In Kolesárová et al. [11] 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 recall the connection between empirical copulas and discrete copulas via permutations. We also make some observations about the permutations that generate the empirical copulas related to associativity and symmetry.

In Section 3 we analyze a new statistic that measures associativity, studying some of its statistical properties under independence in terms of permutations. We also study an auxiliary statistic of associativity which allows to give asymptotic results for the proposed statistic.

In Section 4 we observe that the associativity statistic is non distribution free by simulating its distribution under different Archimedean families. We finally include some general comments about associativity for samples and some remarks connecting the proposed statistic with Spearman's ρ.

2. EMPIRICAL COPULAS, DISCRETE COPULAS AND PERMUTATIONS

We will first recall the definition of the *empirical copula*, see for example Nelsen [17], which is based on classical empirical distribution functions as in 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

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

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

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

where the sample is given by $(1, \sigma(1) = Y_1), (2, \sigma(2) = Y_2), \ldots, (n, \sigma(n) = Y_n)$ and $(\sigma(1),\ \sigma(2), \ldots, \sigma(n))$ is a permutation σ of $\{1, 2, \ldots, n\}$. We define $C'_n(i, 0) = C'_n(0, j) = 0$ for every $i, j \in L$. 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 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. Since 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. [14].

Using the fact that samples that generate discrete copulas C_n are associative if and only if C_n is an ordinal sum of Łukasiewicz matrices as proved in Mayor et al. [14], and the previous statement about symmetry we can construct examples of symmetric samples which are non-associative.

Let us consider the modified sample

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

Then, we have that the discrete copula C'_n equivalent to the empirical copula is given by

$$C'_n(i, j) = \begin{cases} \min\{i, j\} - 1 & \text{if} \quad 1 \leq i, j < n \\ \min\{i, j\} & \text{if} \quad i = n \text{ or } j = n. \end{cases}$$

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.

See Figure 1, where we show a graph of the sample $\underline{X}$ given above.

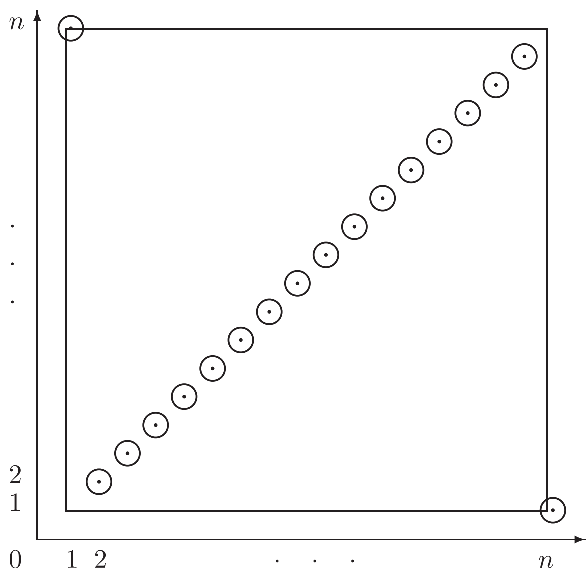

Fig. 1. Symmetric sample $\underline{X_n}$ which is non-associative.

Since associative discrete copulas are ordinal sums of Łukasiewicz permutation matrices, as noticed in Mayor et al. [14] and Kolesárová and Mordelová [12]. In general, associativity has no easy interpretations as noticed in Schweizer and Sklar [19] "The geometric interpretations of the conditions other than associativity are evident; associativity, on the other hand, seems to have no simple geometric interpretation". However, recently Jenei gives a nice geometric interpretation of associativity in [7] and related topics of associativity in [6]. Also, Nelsen [17] mentions "While there does not seem to be a statistical interpretation for random variables with an associative copula, associativity will be a useful property when we construct multivariate Archimedean copulas." In Erdely et al. [4] we observed that we have a very simple and nice geometric representation of associativity in terms of an 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.

In the following section we will define an statistic A^n_π of associativity based on the fact that associative discrete copulas are ordinal sums of Łukasiewicz permutation matrices. We will analyze in detail some of its statistical properties under the hypothesis of independence, and we will find its asymptotic behavior, based on a simpler version $\hat{A}_\pi^n$ of the associativity statistic A_π^n, which we will see that is closely related to it. Specifically we will see that both statistics have the same distribution, at least asymptotically. We will also find the least associative samples.

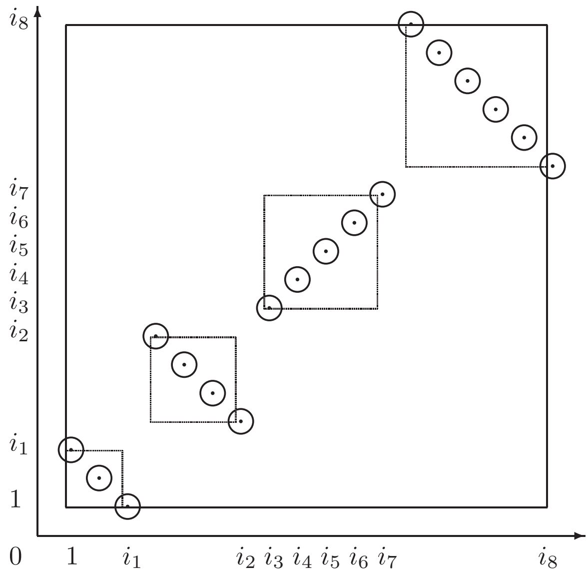

Fig. 2. Associative sample with idempotent elements $i_0 = 0 < i_1 < \cdots < i_8 = n = 18$.

3. A NEW ASSOCIATIVITY STATISTIC OF A COPULA

We have already seen that a modified sample of the form $(1, \sigma(1)), (2, \sigma(2)), \ldots$ $\ldots, (n, \sigma(n))$, where σ is a permutation on I_n, is associative if and only its discrete copula is an ordinal sum of Łukasiewicz permutation matrices, see Figure 2. Let us assume that $1 \le i_1 < i_2 < \cdots < i_k = n$ are the idempotent elements of the sample $(1, \sigma(1)), \ldots, (n, \sigma(n))$, that is $C'_n(i_j, i_j) = i_j$ for $j = 1, 2, \ldots, k$. Observe that n is always an idempotent element of any sample since $C'_n(n, n) = n$. Now we prove an easy result that relates the values of the permutation to the idempotent elements of the sample.

Proposition 3.1. Let $(1, \sigma(1)), (2, \sigma(2)), \ldots, (n, \sigma(n))$ be the modified sample. Then the idempotent elements of the sample are $1 \le i_1 < i_2 < \cdots < i_k = n$ if and only if

$$\{1, 2, \ldots, i_1\} = \{\sigma(1), \sigma(2), \ldots, \sigma(i_1)\},$$

$$\{i_1 + 1, i_1 + 2, \ldots, i_2\} = \{\sigma(i_1 + 1), \sigma(i_1 + 2), \ldots, \sigma(i_2)\},$$

$$\vdots$$

$$\{i_{k-1}+1, i_{k-1}+2, \ldots, n = i_k\} = \{\sigma(i_{k-1}+1), \sigma(i_{k-1}+2), \ldots, \sigma(n) = \sigma(i_k)\},$$

besides for any $1 \le j < i_1$, $\{1, 2, \ldots, j\} \ne \{\sigma(1), \sigma(2), \ldots, \sigma(j)\}$, and for any $1 \le l \le k-1$ and for any $i_l+1 \le j < i_{l+1}$, $\{i_l+1, i_l+2, \ldots, j\} \ne \{\sigma(i_l+1), \sigma(i_l+2), \ldots, \sigma(j)\}$.

Proof. Let $(1, \sigma(1)), (2, \sigma(2)), \ldots, (n, \sigma(n))$ be a modified sample. Let us observe that $C'_n(i, i) = i$ if and only if there are i sample points in the set $\{1, 2, \ldots, i\} \times \{1, 2, \ldots, i\}$, since for each $1 \le i \le n$ there exists a unique $1 \le \sigma(i) \le n$ such that the point $(i, \sigma(i))$ belongs to the sample. Then i is an idempotent of the sample if and only if $C'_n(i, i) = i$ if and only if $\{1, 2, \ldots, i\} = \{\sigma(1), \sigma(2), \ldots, \sigma(i)\}$. In particular n is always an idempotent of the sample.

Therefore $1 \le i_1 < i_2 < \cdots < i_{k-1} < i_k = n$ are all the idempotents of the sample if and only if

$$\{1, 2, \ldots, i_1\} = \{\sigma(1), \sigma(2), \ldots, \sigma(i_1)\},$$

$$\{1, 2, \ldots, i_2\} = \{\sigma(1), \sigma(2), \ldots, \sigma(i_2)\}, \cdots$$

$$\{1, 2, \ldots, i_{k-1}\} = \{\sigma(1), \sigma(2), \ldots, \sigma(i_{k-1})\}$$

and

$$\{1, 2, \ldots, i_k = n\} = \{\sigma(1), \sigma(2), \ldots, \sigma(i_k) = n\}.$$

From these identities it follows also that

$$\{i_1+1, i_1+2, \ldots, i_2\} = \{\sigma(i_1+1), \sigma(i_1+2), \ldots, \sigma(i_2)\},$$

$$\{i_2+1, i_2+2, \ldots, i_3\} = \{\sigma(i_2+1), \sigma(i_2+2), \ldots, \sigma(i_3)\},$$

etc., and

$$\{i_{k-1}+1, i_{k-1}+2, \ldots, n = i_k\} = \{\sigma(i_{k-1}+1), \sigma(i_{k-1}+2), \ldots, \sigma(n) = \sigma(i_k)\}.$$

Observe finally that for any other $j \in \{1, 2, \ldots, n\} \backslash \{i_1, i_2, \ldots i_k = n\}$ we have that

$$\{1, 2, \ldots, j\} \ne \{\sigma(1), \sigma(2), \ldots, \sigma(j)\},$$

which finishes the proof. □

Let us denote by K the number of idempotent elements of a modified sample $(1, \sigma(1)), (2, \sigma(2)), \ldots, (n, \sigma(n))$ of size n. Then $1 \le K \le n$ and K may take all the values between 1 and n. To see this just observe that for n fixed, and using Proposition 3.1, if $\{1, 2, \ldots, i\} \ne \{\sigma(1), \sigma(2), \ldots, \sigma(i)\}$ for any $1 \le i \le n-1$, then $K = 1$, with n the only idempotent element of the sample. Now if $K > 1$, and we assume that $\sigma(1) = 1, \ldots \sigma(k-1) = k-1$, but $\{k, k+1, \ldots, j\} \ne \{\sigma(k), \sigma(k+1), \ldots, \sigma(j)\}$ for any $k \le j < n$, then the sample has k idempotents given by $i_1 = 1, i_2 = 2, \ldots i_{k-1} = k-1$ and $i_k = n$.

Now we define an associativity statistic based on the modified sample $(1, \sigma(1)), \ldots, (n, \sigma(n))$. In order to do so recall that a modified sample is associative if and only if its discrete copula is the ordinal sums of Łukasiewicz permutation matrices, where the number of terms in the sum is determined by the idempotent elements of the sample.

Definition 3.2. Let $(1, \sigma(1)), \ldots, (n, \sigma(n))$ be the modified sample with idempotent elements $1 \leq i_1 < i_2 < \cdots < i_{k-1} < i_k = n$. We define an *associativity statistic* of the sample by:

$$A_\pi^n = \sum_{l=1}^{k} \sum_{j=i_{l-1}+1}^{i_l} (i_l + i_{l-1} - j + 1 - \sigma(j))^2, \tag{2}$$

where $i_0 = 0$. The symbol π is simply to denote the fact that the statistic is based on permutations.

Then A_π^n measures the square distances between the modified sample and the ordinal sum of Łukasiewicz permutation matrices with idempotent elements $1 \leq i_n < i_2 < \cdots < i_{k-1} < i_k = n$. Therefore A_π^n is a measure of the associativity of the modified sample.

For example if σ is such that $\sigma(j) = j$ for $j \in I_n$, then $i_j = j$ for $j = 1, \ldots, n$ and

$$A_\pi^n = \sum_{l=1}^{n} \sum_{j=l-1+1}^{l} (l + (l-1) - j + 1 - \sigma(j))^2 = 0.$$

Therefore, the sample $(1, 1), (2, 2), \ldots, (n, n)$ is associative.

It is important to notice that K the number of idempotent elements in the sample is a random variable which depends on the joint distribution of the vector (X, Y). In the case that $K = 1$, then $i_1 = n$ is the only idempotent of the sample and

$$A_\pi^n = \sum_{j=1}^{n} (n - j + 1 - \sigma(j))^2 = \hat{A}_\pi^n,$$

which is a simpler expression for the associativity statistic, and it will be studied in detail later on. In the following proposition we will find the number of permutations σ of I_n such that $K = 1$ and $K = n$, that is the extreme values of K.

Proposition 3.3. Let $(1, \sigma(1)), \ldots, (n, \sigma(n))$ be the modified sample of size n, where σ is a permutation of I_n. Let us denote by $K_1(j)$ the number of permutations of a modified sample of size j, with only one idempotent. Then a recursive formula for the value of $K_1(n)$ is given by

$$K_1(n) = n! - \sum_{j=1}^{n-1} K_1(j)(n-j)!, \tag{3}$$

where $K_1(1) = 1$. Besides there is only one permutation σ of I_n such that $K = n$.

Proof. Let us assume that $n = 1$, then the only permutation of I_1 is $\sigma(1) = 1$, in this case the only idempotent is $i_1 = 1$, and $K_1(1) = 1$. If $n = 2$, then we have two permutations of I_2. If σ is such that $\sigma(1) = 1$ and $\sigma(2) = 2$, then the idempotents of the sample are $i_1 = 1 < i_2 = 2$, and the number of idempotents is $K = 2$. On the

the other hand, if $\sigma(1) = 2$ and $\sigma(2) = 1$, then the only idempotent is $i_1 = 2$ and $K = 1$. Observe that in this case

$$K_1(2) = 2! - K_1(1)(2-1)! = 1.$$

Therefore the formula (3) holds for $n = 2$.

Now the proof follows by induction. Let us assume that the formula (3) holds for $n-1$. We have to prove that it also holds for n. We will first count the number of permutations σ of I_n such that the number of idempotents of the sample generated by these permutations is greater than one, that is $K > 1$. In order to see this, we just have to observe the position of the first idempotent i_1.

Now, $i_1 = 1$ if and only if $\sigma(1) = 1$. So, if $\sigma(1) = 1$, then $K > 1$, and the remaining values of $\sigma(2), \ldots, \sigma(n)$ have no restrictions. Observe that if we ask that $\sigma(1) = 1$, it is equivalent to ask that a sample of size one has only one idempotent, which happens only in one way. Therefore, if $\sigma(1) = 1$, there are $K_1(1)(n-1)!$ permutations σ of I_n such that $K > 1$.

Now assume that $i_1 = k$, for $2 \leq k < n-1$. By Proposition 3.1, in order to get a permutation that satisfies the previous condition is necessary to ask that

$$\{1, 2, \ldots, k\} = \{\sigma(1), \sigma(2), \ldots, \sigma(k)\},$$

but for any $1 \leq j < k$, $\{1, 2, \ldots, j\} \neq \{\sigma(1), \sigma(2), \ldots, \sigma(j)\}$. Observe that in I_k, the number of permutations that satisfy this condition is exactly $K_1(k)$. Now if $i_1 = k$, then the values of $\sigma(k+1), \ldots, \sigma(n)$ have no restrictions. Therefore, if $i_1 = k$, there are $K_1(k)(n-k)!$ permutations such that $K > 1$.

Adding all the cases in which $K > 1$ depending on the value of i_1 such that $i_1 < n$. We obtain that the number of permutations σ of I_n such that $K > 1$ is given by:

$$\sum_{j=1}^{n-1} K_1(j)(n-j)!.$$

And since there are $n!$ possible permutations σ of I_n, then

$$K_1(n) = n! - \sum_{j=1}^{n-1} K_1(j)(n-j)!.$$

Finally observe that $K = n$ if and only if every $i \in I_n$ is an idempotent element of the sample, this happens if and only if $\sigma(i) = i$ for every $i \in I_n$ according to Proposition 3.1. Therefore, there exists only one permutation σ of I_n such that $K = n$. $\square$

Now, let $\underline{X_n} = (X_1, Y_1),\ (X_2, Y_2),\ \ldots, (X_n, Y_n)$ be a bivariate random sample of size n where X and Y are continuous random variables with copula C. Let $(1, \sigma(1)), (2, \sigma(2)), \ldots, (n, \sigma(n))$ be the modified sample induced by the original sample. In this case σ the permutation and K the number of idempotent elements of the discrete copula C_n' are random. In particular, when $C = \Pi$ is the product copula we have immediately from Proposition 3.3 the following:

Corollary 3.4. 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 continuous and independent 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 and let K be the number of idempotent elements of the discrete copula C_n induced by the modified sample. Then

$$\mathrm{P}\,(K = n) = \frac{1}{n!} \quad \text{and} \quad \mathrm{P}\,(K = 1) = 1 - \frac{\sum_{j=1}^{n-1} K_1(j)(n-j)!}{n!}.$$

In Table 1 we give the values of $\mathrm{P}\,(K = 1)$ for different values of n the sample size, as can be observed the value of $\mathrm{P}\,(K = 1)$ for $n \geq 2$ increases to one as n increases, and also $\mathrm{P}\,(K = 1) \approx 1 - 2/n$ for $n \geq 2$.

Table 1. $\mathrm{P}\,(K = 1)$ for different values of n and values of $1 - 2/n$.

value of n	$\mathrm{P}\,(K = 1)$	$1 - 2/n$
1	1	–
2	0.5	0
3	0.5	0.3333
4	0.5416	0.5
5	0.5916	0.6
6	0.6402	0.666
7	0.6839	0.714
8	0.7217	0.750
9	0.7532	0.777
10	0.7796	0.8
50	0.9595	0.96
100	0.9798	0.98
200	0.989974	0.99
300	0.993322	0.9933
400	0.994994	0.995
500	0.995956	0.996
1000	0.997999	0.998

Let us study some properties of the associativity statistic defined in equation (2). First we will prove that in the case of the existence of more than one idempotent, A_π^n is a sum of variables of the same type in dimensions smaller than or equal n.

Lemma 3.5. Let $(1, \sigma(1)), (2, \sigma(2)), \ldots, (n, \sigma(n))$ be the modified sample of size n and assume that $1 \leq i_1 < i_2 \cdots < i_k = n$ are its idempotents. Then for each $l = 1, 2, \ldots, k$ there exists a unique permutation σ_l of $\{1, 2, \ldots, i_l - i_{l-1}\}$, where

$i_0 = 0$, such that

$$\sum_{j=i_{l-1}+1}^{i_l} (i_l + i_{l-1} - j + 1 - \sigma(j))^2 = \sum_{k=1}^{i_l - i_{l-1}} (i_l - i_{l-1} - k + 1 - \sigma_l(k))^2.$$

P r o o f. If $l = 1$ since $i_{l-1} = i_0 = 0$ and i_1 is the first idempotent, taking $\sigma'(k) = \sigma(k)$ for $k = 1, 2, \ldots, i_1$ the result follows.

If $1 < l \le k$, since i_{l-1} and i_l are consecutive idempotents, from Proposition 3.1 we know that

$$\{i_{l-1}+1, i_{l-1}+2, \ldots, i_l\} = \{\sigma(i_{l-1}+1), \sigma(i_{l-1}+2), \ldots, \sigma(i_l)\}.$$

Let us define

$$\sigma'(k) = \sigma(i_{l-1}+k) - i_{l-1} \quad \text{for} \quad k = 1, 2, \ldots, i_l - i_{l-1}.$$

Then σ' is a unique permutation of $\{1, 2, \ldots, i_l - i_{l-1}\}$, and

$$\begin{aligned}
\sum_{k=1}^{i_l - i_{l-1}} (i_l - i_{l-1} - k + 1 - \sigma'(k))^2 &= \sum_{k=1}^{i_l - i_{l-1}} (i_l - i_{l-1} - k + 1 - (\sigma(i_{l-1}+k) - i_{l-1}))^2 \\
&= \sum_{k=1}^{i_l - i_{l-1}} (i_l - k + 1 - \sigma(i_{l-1}+k))^2 \\
&= \sum_{j=i_{l-1}+1}^{i_l} (i_l + i_{l-1} - j + 1 - \sigma(j))^2,
\end{aligned}$$

where $j = i_{l-1} + k$. $\square$

Let us assume again that $\underline{X_n} = (X_1, Y_1), (X_2, Y_2), \ldots, (X_n, Y_n)$ is a bivariate random sample of size n where X and Y are continuous random variables with copula C. Let $(1, \sigma(1)), (2, \sigma(2)), \ldots, (n, \sigma(n))$ be the modified sample generated by $\underline{X_n}$. Let us define an auxiliary event by

$$J_{i_1 i_2 \cdots i_k} = \{\text{the idempotents associated to } \sigma \text{ are } 1 \le i_1 < i_2 \cdots < i_k = n\} \tag{4}$$

for some $1 \le k \le n$, with $i_k = n$. For example J_n is the event that the only idempotent element associated to σ is $i_1 = n$, and in this case $K = 1$. Conditioning on these events we can find an expression for the density of the random variable A_π^n, as shown in the following:

Theorem 3.6. 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 continuous and independent 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 and assume that $1 \le i_1 < i_2 < \cdots i_k = n$ are the idempotents associated to σ and $i_0 = 0$. Then

$$\mathrm{P}\left(J_{i_1 i_2 \cdots i_k}\right) = \frac{K_1(i_1)K_1(i_2 - i_1)\cdots K_1(n - i_{k-1})}{n!}, \tag{5}$$

where $K_1(l)$ is the number of permutations τ of $\{1, 2, \ldots, l\}$ with only one idempotent, which is necessarily l. Besides, for any nonnegative integer a

$$\mathrm{P}\left(A_\pi^n = a\right) = \sum_{k=1}^{n} \sum_{\underline{i} \in R_k} \sum_{\underline{b} \in B_a} \left\{\Pi_{j=1}^{k} P\left(A_\pi^{i_j - i_{j-1}} = b_j | J_{i_1 i_2 \cdots i_k}\right)\right\} \mathrm{P}\left(J_{i_1 i_2 \cdots i_k}\right), \tag{6}$$

where

$$R_k = \{\underline{i} = (i_1, i_2, \ldots, i_k) \mid 1 \le i_1 < i_2 < \cdots i_{k-1} < i_k = n \text{ are } k \text{ integers}\},$$

$$B_a = \{\underline{b} = (b_1, b_2, \ldots, b_k) \mid b_1, b_2, \ldots, b_k \ge 0 \text{ are integers, and } \sum_{j=1}^{k} b_j = a\}. \tag{7}$$

P r o o f. Let $(1, \sigma(1)), (2, \sigma(2)), \ldots (n, \sigma(n))$ be a modified sample of X and Y two continuous and independent random variables. Let us assume that $1 \le i_1 < i_2 < \cdots < i_k = n$ are the idempotents associated to the permutation σ of I_n. Using Proposition 3.1 we know that

$$\{i_{l-1} + 1, i_{l-1} + 2, \ldots, i_l\} = \{\sigma(i_{l-1} + 1), \sigma(i_{l-1} + 2), \ldots, \sigma(i_l)\},$$

for each $l = 1, 2, \ldots, k$, and also for any $j \in \{1, 2, \ldots, i_1 - i_{l-1} - 1\}$

$$\{i_{l-1} + 1, i_{l-1} + 2, \ldots, i_{l-1} + j\} \ne \{\sigma(i_{l-1} + 1), \sigma(i_{l-1} + 2), \ldots, \sigma(i_{l-1} + j)\}.$$

If we consider the function that maps $i_{l-1} + s$ into s, for $s = 1, 2, \ldots, i_l - i_{l-1}$ and we define the permutation σ' of $\{1, 2, \ldots, i_l - i_{l-1}\}$ such that $\sigma'(s) = \sigma(i_{l-1} + s) - i_{l-1}$ for $s = 1, 2, \ldots, i_l - i_{l-1}$. Then we have that

$$\{1, 2, \ldots, i_l - i_{l-1}\} = \{\sigma'(1), \sigma'(2), \ldots, \sigma'(i_l - i_{l-1})\}$$

but for any $j \in \{1, 2, \ldots, i_l - i_{l-1} - 1\}$

$$\{1, 2, \ldots, j\} \ne \{\sigma'(1), \sigma'(2), \ldots, \sigma'(j)\}.$$

Therefore using Proposition 3.3 there are exactly $K_1(i_l - i_{l-1})$ permutations σ' of $\{1, 2, \ldots, i_l - i_{l-1}\}$ that satisfy the above conditions. Now by letting l vary from 1 to k we obtain that

$$\mathrm{P}\left(J_{i_1 i_2 \cdots i_k}\right) = \frac{\Pi_{l=1}^{k} K_1(i_l - i_{l-1})}{n!},$$

which proves equation (5).

Recall that the associativity statistic is defined by

$$A_\pi^n = \sum_{l=1}^{k} \sum_{j=i_{l-1}+1}^{i_l} (i_l + i_{l-1} - j + 1 - \sigma(j))^2.$$

Then conditioning on the idempotents, we have that for any a nonnegative integer

$$\mathrm{P}\,(A_\pi^n = a) = \sum_{k=1}^{n} \sum_{\underline{i} \in R_k} \mathrm{P}\,(A_\pi^n = a | J_{i_1 i_2 \cdots i_k}) \mathrm{P}\,(J_{i_1 i_2 \cdots i_k}).$$

Now observe that using Lemma 3.5 A_π^n can be rewritten as

$$A_\pi^n = \sum_{l=1}^{k} \sum_{j=1}^{i_l - i_{l-1}} (i_l - i_{l-1} - j + 1 - \sigma_l'(j))^2,$$

where σ_l' is a permutation of the set $\{1, 2, \ldots, i_l - i_{l-1}\}$ for every $1 \leq l \leq k$. If we let

$$A_\pi^{i_l - i_{l-1}} = \sum_{j=1}^{i_l - i_{l-1}} (i_l - i_{l-1} - j + 1 - \sigma_l'(j))^2 \text{ for every } 1 \leq l \leq k.$$

Then

$$A_\pi^n = \sum_{l=1}^{k} A_\pi^{i_l - i_{l-1}}. \tag{8}$$

Hence, for any a nonegative integer

$$\mathrm{P}\,(A_\pi^n = a) = \sum_{k=1}^{n} \sum_{\underline{i} \in R_k} P\left(A_\pi^{i_1} + A_\pi^{i_2 - i_1} + \cdots + A_\pi^{n - i_{k-1}} = a | J_{i_1 i_2 \cdots i_k}\right) \mathrm{P}\,(J_{i_1 i_2 \cdots i_k})$$

$$= \sum_{k=1}^{n} \sum_{\underline{i} \in R_k} \sum_{\underline{b} \in B_a} P\left(A_\pi^{i_1} = b_1, A_\pi^{i_2 - i_1} = b_2, \ldots, A_\pi^{n - i_{k-1}} = b_k | J_{i_1 i_2 \cdots i_k}\right) \mathrm{P}\,(J_{i_1 i_2 \cdots i_k}).$$

Now finally observe that $A_\pi^{i_1}, A_\pi^{i_2 - i_1}, \ldots, A_\pi^{n - i_{k-1}}$ are conditionally independent given $J_{i_1 i_2 \cdots i_k}$, since the permutations σ_l' only depend on the length $i_l - i_{l-1}$. Therefore,

$$\mathrm{P}\,(A_\pi^n = a) = \sum_{k=1}^{n} \sum_{\underline{i} \in R_k} \sum_{\underline{b} \in B_a} \left\{\Pi_{j=1}^{k} P\left(A_\pi^{i_j - i_{j-1}} = b_j | J_{i_1 i_2 \cdots i_k}\right)\right\} \mathrm{P}\,(J_{i_1 i_2 \cdots i_k}),$$

as we wanted to prove. □

Observe that from Theorem 3.6, if σ is a permutation of I_n with more than one idempotent and we know the distribution of A_π^j, for any $1 \leq j \leq n-1$. Then we can find inductively $\mathrm{P}\,(A_\pi^n = a | J_{i_1 i_2 \cdots i_k})$ for any $a \geq 0$ and $\underline{i} \in R_k$. The only cases in which we have to find $\mathrm{P}\,(A_\pi^n = a | J_{i_1 i_2 \cdots i_k})$ explicitly is when σ induces only one idempotent, that is, when $i_1 = n$ and $K = 1$. That is, we have to evaluate all the conditional probabilities of the form $\mathrm{P}\left(A_\pi^n = a | J_n\right)$. The exact number of evaluations that we have to do is $K_1(n)$ given in Proposition 3.3.

As an easy example we provide the actual density of A_π^5 which is given by

$$P\left(A_\pi^5 = a\right) = \begin{cases} 16/120 & \text{if} \quad a = 0 \\ 20/120 & \text{if} \quad a = 2 \\ 5/120 & \text{if} \quad a = 4 \\ 14/120 & \text{if} \quad a = 6 \\ 11/120 & \text{if} \quad a = 8 \\ 10/120 & \text{if} \quad a = 10, 14, 18 \\ 4/120 & \text{if} \quad a = 12, 20, 22, 24 \\ 6/120 & \text{if} \quad a = 16 \\ 2/120 & \text{if} \quad a = 26. \end{cases}$$

In Figure 3, we graph the density of A_π^5. Observe that this density is quite asymmetric, and it has several modes. This behavior also holds for larger values of n. An important question which we will solve next is: What is the actual range of A_π^n? We will need to answer this question to analyze in detail the distribution of the statistic A_π^n.

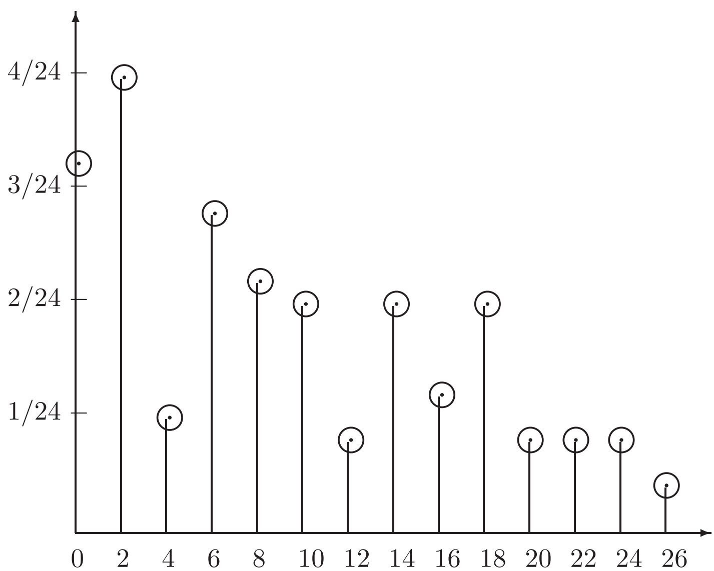

Fig. 3. Density of A_π^5.

Now we will see which is the range of A_π^n and we will find the probability of its extreme values.

Lemma 3.7. 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 continuous and independent 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 and assume that $1 \le i_1 < i_2 < \cdots i_k = n$ are the idempotents associated to σ and $i_0 = 0$. Then the range of A_π^n, denoted by $\mathrm{Ran}(A_\pi^n)$, is given by

$$\mathrm{Ran}(A_\pi^n) = \left\{0, 2, 4, \ldots, \frac{n^3 - 13n + 18}{3}\right\}.$$

Besides,

$$\mathrm{P}\left(A_\pi^n = 0\right) = \frac{2^{n-1}}{n!} \quad \text{and} \quad P\left(A_\pi^n = \frac{n^3 - 13n + 18}{3}\right) = \frac{2}{n!}, \text{ if } n \ge 3,$$

and

$$\mathrm{P}\left(A_\pi^1 = 0\right) = \mathrm{P}\left(A_\pi^2 = 0\right) = 1.$$

Proof. Using equation (8) in the proof of Theorem 3.6, we have that

$$A_\pi^n = \sum_{l=1}^{k} A_\pi^{i_l - i_{l-1}} = \sum_{l=1}^{k} \sum_{j=1}^{i_l - i_{l-1}} (i_l - i_{l-1} - j + 1 - \sigma_l'(j))^2.$$

In order to see that A_π^n only takes values on the set of nonnegative even integers, just observe that for any $j \ge 1$ and τ, σ permutations in the set $\{1, 2, \ldots, j\}$ we have that $\sum_{k=1}^{j} (\tau(k) - \sigma(k))^2$ is a nonnegative even integer, since

$$\begin{aligned}
\sum_{k=1}^{j} (\tau(k) - \sigma(k))^2 &= \sum_{k=1}^{j} \tau(k)^2 - 2\sum_{k=1}^{j} \tau(k)\sigma(j) + \sum_{k=1}^{j} \sigma(k)^2 \\
&= 2\left(\frac{j(j+1)(2j+1)}{6} - \sum_{k=1}^{j} \tau(j)\sigma(j)\right),
\end{aligned}$$

since σ and τ are bijections from I_j onto I_j.

Now observe that

$$\begin{aligned}
\sum_{i=1}^{n} (n - i + 1 - i)^2 &= \sum_{i=1}^{n} (n + 1 - 2i)^2 \\
&= \sum_{i=1}^{n} \left((n+1)^2 - 4(n+1)i + 4i^2\right) \\
&= n(n+1)^2 - 2n(n+1)^2 + \frac{4}{6} n(n+1)(2n+1) \\
&= n(n+1)\left(\frac{n-1}{3}\right) \\
&= \frac{n^3 - n}{3}. \qquad (9)
\end{aligned}$$

This is clearly the maximum value for $\sum_{i=1}^{n}(n-i+1-\sigma(i))^2$ when we vary σ over all permutations of I_n. Observe that the terms $(n-1)^2$ and $(1-n)^2 = (n-1)^2$ appear in this sum. In fact it is clear that

$$\sum_{i=1}^{n}(n-i+1-i)^2 = 2\sum_{k=0}^{[n/2]-1}(n-(2k+1))^2, \tag{10}$$

where $[n/2]$ is the greatest integer less or equal $n/2$.

However, $(n^3-n)/3$ is not a possible value for A_π^n, since if $\sigma(i) = i$ for $i = 1, 2, \ldots, n$, then in fact $A_\pi^n = 0$, as noticed after Definition 3.2 (every i from 1 to n are idempotents). The maximum of A_π^n must be attained when there is only one idempotent, that is , when $K = 1$. Let us see that the value $(n-1)^2$ is not a term in $\sum_{i=1}^{n}(n-i+1-\sigma(i))^2$, if the only idempotent is $i_1 = n$. The only possibilities in order to have that the term $(n-1)^2$ appears in the previous sum are if $\sigma(1) = 1$ or $\sigma(n) = n$, but in both cases there are at least two idempotents, respectively 1 and n, or $n-1$ and n.

Nevertheless, the term $(n-2)^2$ is always possible in $\sum_{i=1}^{n}(n-i+1-\sigma(i))^2$, when $i_1 = n$ is the only idempotent of σ.

From equation (10) we will complete a permutation σ of I_n with one idempotent, such that

$$\sum_{i=1}^{n}(n-i+1-\sigma(i))^2 = 2(n-2)^2 + 2\sum_{k=1}^{[n/2]-1}(n-(2k+1))^2. \tag{11}$$

Let us assume that n is an odd integer of the form $n = 2l+1$. We will ask that the permutation satisfies the following condition:

$$\text{If } \sigma(k)=s \text{ then } \sigma(n-\sigma(k)+1)=\sigma(n-s+1)=n-k+1, \text{ for every } k\in I_n. \tag{12}$$

First, let us define $\sigma(1) = 2$ then using (12), $\sigma(n-\sigma(1)+1) = \sigma(n-1) = n$. Then $(n-1+1-\sigma(1))^2 = (n-2)^2$ and $(n-(n-1)+1-\sigma(n-1))^2 = (2-n)^2$.

Now, for $k = 2j$ and $1 \le j < [n/2] = l$ define $\sigma(2j) = 2j+2$, then from equation (12), $\sigma(n-\sigma(2j)+1) = \sigma(n-2j-1) = n-2j+1$. Then $(n-2j+1-\sigma(2j))^2 = (n-2j+1-(2j+2))^2 = (n-4j-1)^2 = (2l-4j)^2$ and $(n-(n-2j-1)+1-\sigma(n-2j-1))^2 = (2j+2-(n-2j+1))^2 = (4j+1-n)^2 = (n-4j-1)^2 = (2l-4j)^2$.

If $k = 2j+1$ and $1 \le j \le [n/2] = l$ define $\sigma(2j+1) = 2j-1$, then from equation (12), $\sigma(n-\sigma(2j+1)+1) = \sigma(n-2j+2) = n-2j$. Then $(n-(2j+1)+1-\sigma(2j+1))^2 = (n-2j-(2j-1))^2 = (n-4j+1)^2 = (2l+2-4j)^2$ and $(n-(n-2j+2)+1-\sigma(n-2j+2))^2 = (2j-1-(n-2j))^2 = (4j-1-n)^2 = (n-4j+1)^2 = (2l+2-4j)^2$.

For example if $n = 11$, and we define the permutation σ in I_{11} described above, that is, $\sigma = (2, 4, 1, 6, 3, 8, 5, 10, 7, 11, 9)$. Then

$$\begin{aligned}\sum_{i=1}^{11}(11-i+1-\sigma(i))^2 &= 9^2+6^2+8^2+2^2+4^2+2^2+0+6^2+4^2+9^2+8^2 \\ &= 2*(11-2)^2 + 2\sum_{k=1}^{4}(11-(2k+1))^2,\end{aligned}$$

then σ satisfies equation (11), and clearly the only idempotent of σ is $i_1 = 11$.

Now assume that n is an even integer of the form $n = 2l$. Let us define σ in I_n such that $\sigma(1) = 2$, $\sigma(n) = n-1$, $\sigma(2j) = 2j+2$, for $1 \leq j < [n/2] = l$, and $\sigma(2j+1) = 2j-1$ for $1 \leq j < [n/2]$. Then, as above, it can be seen that σ satisfies equation (11).

For example if $n = 10$, and we define the permutation σ in I_{10} described above, that is, $\sigma = (2, 4, 1, 6, 3, 8, 5, 10, 7, 9)$. Then

$$\begin{aligned}\sum_{i=1}^{10}(10-i+1-\sigma(i))^2 &= 8^2+5^2+7^2+1^2+3^2+3^2+1^2+7^2+5^2+8^2 \\ &= 2*(10-2)^2+2\sum_{k=1}^{4}(10-(2k+1))^2,\end{aligned}$$

then σ satisfies equation (11), and clearly the only idempotent of σ is $i_1 = 10$.

Now using equations (9), (10) and (11) we can easily obtain the maximum value M_n of A_π^n, since

$$\begin{aligned}M_n &= \sum_{i=1}^{n}(n-i+1-\sigma(i))^2 - 2(n-1)^2 + 2(n-2)^2 \\ &= \frac{n^3-n-6(n-1)^2+6(n-2)^2}{3} \\ &= \frac{n^3-13n+18}{3}.\end{aligned}$$

It can be seen that A_π^n takes any even value between 0 and M_n. Therefore, $\mathrm{Ran}(A_\pi^n) = \{0, 2, \ldots, M_n\}$.

For each $n \geq 1$ there exists just another similar construction of a permutation σ of I_n, such that $\sigma(1) = 3$, σ has only one idempotent, and σ satisfies equation (11). To see this we just have to find the symmetric permutation of the above case, for example for $n = 10$ we observed that the permutation $\sigma = (2, 4, 1, 6, 3, 8, 5, 10, 7, 9)$ leads to the maximum value of A_π^{10}, that is, M_{10}. Then finding the symmetric permutation of σ, that is, $\sigma' = (3, 1, 5, 2, 7, 4, 9, 6, 10, 8)$, observe that $\sigma(\sigma'(i)) = i$ for $i = 1, 2, \ldots, 10$. This case also leads to the maximum value of A_π^{10}. Therefore, $\mathrm{P}\left(A_\pi^n = M_n\right) = 2/n!$.

Finally, if σ is a permutation of I_n, with k idempotents, for $1 \leq k \leq n$, since $i_k = n$ is always an idempotent, then the remaining $k-1$ idempotents can be selected in

$$\binom{n-1}{k-1}$$

ways, and for each selection of $1 \leq i_1 < i_2 < \cdots < i_{k-1} < i_k = n$ of idempotents of σ there is only one permutation σ such that $A_\pi^n = 0$. Therefore, there are

$$\sum_{k=0}^{n-1}\binom{n-1}{k-1} = 2^{n-1}$$

cases such that $A_\pi^n = 0$. Kolesárová and Mordelová [12] were the first to observe that there are 2^{n-1} associative discrete copulas on L. Hence, $\mathrm{P}\,(A_\pi^n = 0) = \frac{2^{n-1}}{n!}$. The last statement of the Lemma follows straightforward. □

Observe that according to Lemma 3.7, equation (7) in Theorem 3.6 can be rewritten as

$$B_a = \left\{ \underline{b} = (b_1, b_2, \ldots, b_k) \,|\, b_1, \ldots b_k \geq 0 \text{ are } \textit{even} \text{ integers, and } \sum_{j=1}^{k} b_j = a \right\},$$

where a is a nonegative *even* integer.

Now we will study a statistic which is close and related to A_π^n defined by

$$\hat{A}_\pi^n = \sum_{j=1}^{n} (n - j + 1 - \sigma(j))^2. \tag{13}$$

If we have a modified sample of the form $(1, \sigma(1)), (2, \sigma(2)), \ldots, (n, \sigma(n))$ and there is only one idempotent associated to the permutation σ of I_n, that is, $K = 1$ then

$$A_\pi^n = \hat{A}_\pi^n \qquad \text{if} \qquad K = 1. \tag{14}$$

Now we will find the expectation and variance of $\hat{A}_\pi^n$ under independence. The results in Proposition 3.8 and 3.9 could be obtained using ranks, see for example Hettmansperger [5], but we include them here for completeness, and the fact that interesting observations can be derived from them.

Proposition 3.8. 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 continuous and independent 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. Then

$$\mathrm{E}\,(\sigma(j)^k) = \frac{\sum_{l=1}^{n} l^k}{n} \quad \text{for every} \quad k \geq 1, \tag{15}$$

$$\mathrm{E}\,(\sigma(j)\sigma(k)) = \frac{(n+1)(3n+2)}{12} \quad \text{for every} \quad 1 \leq j \neq k \leq n. \tag{16}$$

Besides, the expectation and variance of $\hat{A}_\pi^n$ are given by

$$\mathrm{E}\,(\hat{A}_\pi^n) = \frac{n(n-1)(n+1)}{6} \tag{17}$$

and

$$\mathrm{Var}(\hat{A}_\pi^n) = \frac{n^2(n+1)^2(n-1)}{36}. \tag{18}$$

P r o o f. Let $(1, \sigma(1)), (2, \sigma(2)), \ldots, (n, \sigma(n))$ a modified sample of two continuous and independent random variables. Then every permutation σ of I_n has probability $1/n!$. Hence, for every $j \in I_n$,

$$\mathrm{P}\left(\sigma(j)=i\right)=\begin{cases}\frac{(n-1)!}{n!}=\frac{1}{n} & \text{if} \quad 1\le i\le n\\ 0 & \text{otherwise.}\end{cases}$$

Besides, if $1 \le j, k \le n$ with $j \neq k$, then

$$\mathrm{P}\left(\sigma(j)=i,\sigma(k)=l\right)=\begin{cases}\frac{(n-2)!}{n!}=\frac{1}{n(n-1)} & \text{if} \quad 1\le i\neq l\le n\\ 0 & \text{otherwise.}\end{cases}$$

Therefore,

$$\mathrm{E}\left(\sigma(j)^k\right)=\sum_{i=1}^{n} i^k \mathrm{P}\left(\sigma(j)=i\right)=\sum_{i=1}^{n}\frac{i^k}{n},$$

if $k \ge 1$, which proves (15). In particular,

$$\mathrm{E}\left(\sigma(j)\right)=\frac{n+1}{2} \quad \text{and} \quad \mathrm{E}\left((\sigma(j))^2\right)=\frac{(n+1)(2n+1)}{6}. \tag{19}$$

Now, if $j, k \in I_n$ with $j \neq k$, then

$$\begin{aligned}\mathrm{E}\left(\sigma(j)\sigma(k)\right) &= \sum_{i=1}^{n}\sum_{l=1,l\neq i}^{n}\frac{il}{n(n-1)}\\ &= \frac{1}{n(n-1)}\sum_{i=1}^{n} i\left(\frac{n(n+1)}{2}-i\right)\\ &= \frac{1}{n(n-1)}\left(\frac{n^2(n+1)^2}{4}-\frac{n(n+1)(2n+1)}{6}\right)\\ &= \frac{(n+1)(3n+2)}{12},\end{aligned}$$

which proves (16). Now since $\tau(j) = n - j + 1$, for $j = 1, 2, \ldots, n$, and σ are permutations of I_n. Then

$$\begin{aligned}\hat{A}^n_\pi &= \sum_{j=1}^{n}(n-j+1-\sigma(j))^2\\ &= \sum_{j=1}^{n}(n-j+1)^2-2\sum_{j=1}^{n}(n+1-j)\sigma(j)+\sum_{j=1}^{n}(\sigma(j))^2\\ &= \frac{n(n+1)(2n+1)}{3}-n(n+1)^2+2\sum_{j=1}^{n} j\sigma(j)\\ &= 2\sum_{j=1}^{n} j\sigma(j)-\frac{n(n+1)(n+2)}{3}.\end{aligned} \tag{20}$$

Now, using equations (19) and (20) we have that

$$\begin{aligned}
\mathrm{E}\left(\hat{A}_\pi^n\right) &= 2\sum_{j=1}^{n} j\mathrm{E}\left(\sigma(j)\right) - \frac{n(n+1)(n+2)}{3} \\
&= \frac{n(n+1)^2}{2} - \frac{n(n+1)(n+2)}{3} \\
&= n(n+1)\left(\frac{n+1}{2} - \frac{n+2}{3}\right) \\
&= \frac{n(n-1)(n+1)}{6},
\end{aligned}$$

proving (17). Now, using equation (20), and letting $C_n = n(n+1)(n+2)/3$,

$$\begin{aligned}
\mathrm{E}\left((\hat{A}_\pi^n)^2\right) &= 4\sum_{j=1}^{n}\sum_{k=1}^{n} jk\mathrm{E}\left(\sigma(j)\sigma(k)\right) - 4C_n\sum_{j=1}^{n} j\mathrm{E}\left(\sigma(j)\right) + C_n^2 \\
&= 4\sum_{j=1}^{n} j^2\mathrm{E}\left((\sigma(j)^2\right) + 4\sum_{j=1}^{n}\sum_{k=1,k\neq j}^{n} jk\mathrm{E}\left(\sigma(j)\sigma(k)\right) \\
&\quad -4C_n\sum_{j=1}^{n} j\mathrm{E}\left(\sigma(j)\right) + C_n^2 \\
&= a_1 + a_2 + a_3 + a_4. \qquad (21)
\end{aligned}$$

From equation (19),

$$a_1 = 4\frac{(n+1)(2n+1)}{6}\frac{n(n+1)(2n+1)}{6} = \frac{n(n+1)^2(2n+1)^2}{9},$$

from equation (16),

$$\begin{aligned}
a_2 &= \frac{4(n+1)(3n+2)}{12}\sum_{j=1}^{n} j\left(\frac{n(n+1)}{2} - j\right) \\
&= \frac{4n(n+1)^2(3n+2)}{24}\frac{n(n+1)}{2} - \frac{4(n+1)(3n+2)}{12}\frac{n(n+1)(2n+1)}{6} \\
&= \frac{n^2(n+1)^3(3n+2)}{12} - \frac{n(n+1)^2(3n+2)(2n+1)}{18}
\end{aligned}$$

$$a_3 = -C_n n(n+1)^2 = -\frac{n^2(n+1)^3(n+2)}{3},$$

and

$$a_4 = C_n^2 = \frac{n^2(n+1)^2(n+2)^2}{9}.$$

Replacing a_1, a_2, a_3 and a_4 in (21) we get

$$\begin{aligned}
\mathrm{E}\,((\hat{A}^n_\pi)^2) &= \frac{n(n+1)^2}{36}\{4(2n+1)^2 + 3n(n+1)(3n+2) \\
&\quad -2(3n+2)(2n+1) - 12n(n+1)(n+2) + 4n(n+2)^2\} \\
&= \frac{n^3(n+1)^2(n-1)}{36}. \qquad (22)
\end{aligned}$$

Finally using equations (17) and (22)

$$\begin{aligned}
\mathrm{Var}(\hat{A}^n_\pi) &= \mathrm{E}\,((\hat{A}^n_\pi)^2) - (\mathrm{E}\,(\hat{A}^n_\pi))^2 \\
&= \frac{n^3(n+1)^2(n-1)}{36} - \frac{n^2(n+1)^2(n-1)^2}{36} \\
&= \frac{n^2(n+1)^2(n-1)}{36}(n-(n-1)) \\
&= \frac{n^2(n+1)^2(n-1)}{36},
\end{aligned}$$

which is equation (18). □

Now we will find some properties of the statistic $\hat{A}^n_\pi$, relating its distribution with any other fixed permutation τ, and using the fact that $\hat{A}^n_\pi$ is the sum of n random variables.

Proposition 3.9. 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 continuous and independent 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. Let τ be a fixed permutation of I_n. Define

$$\tilde{A}^n_\pi = \sum_{j=1}^{n} (\tau(j) - \sigma(j))^2, \text{ and recall that } \hat{A}^n_\pi = \sum_{j=1}^{n} (n - j + 1 - \sigma(j))^2.$$

Then

$$\tilde{A}^n_\pi \stackrel{d}{=} \hat{A}^n_\pi, \qquad (23)$$

where $\stackrel{d}{=}$ stands for "equals in distribution". Let $Y_i = 2(R_n - i\sigma(i))$, for $i = 1, 2, \ldots, n$, where $R_n = (n+1)(2n+1)/6$. Then the covariances $\mathrm{Cov}(Y_i, Y_j)$ and correlations $\rho(Y_i, Y_j)$ are given by:

$$\mathrm{Cov}(Y_i, Y_j) = \frac{-(n+1)}{3} ij \quad \text{and} \quad \rho(Y_i, Y_j) = \frac{-1}{(n-1)}, \qquad (24)$$

for every $i, j \in \{1, 2, \ldots, n\}$ with $i \neq j$.

Besides, $\hat{A}^n_\pi$ is a symmetric random variable with respect to $(n^3 - n)/6$.

P r o o f. First let us observe that defining $\tau(j) = n - j + 1$ for $j = 1, 2, \ldots, n$, we obtain a permutation τ of I_n. So, $\hat{A}_\pi^n$ is of the form $\sum_{j=1}^n (\tau(j) - \sigma(j))^2$ for a fixed permutation τ of I_n. Now, let τ be a fixed permutation of I_n, then

$$\begin{aligned}
\tilde{A}_\pi^n &= \sum_{j=1}^n (\tau(j) - \sigma(j))^2 \\
&= \sum_{j=1}^n \tau(j)^2 - 2\sum_{j=1}^n \tau(j)\sigma(j) + \sum_{j=1}^n \sigma(j)^2 \\
&= \frac{2n(n+1)(2n+1)}{6} - 2\sum_{j=1}^n \tau(j)\sigma(j) \\
&= \frac{2n(n+1)(2n+1)}{6} - 2\sum_{k=1}^n k\sigma'(k), \qquad (25)
\end{aligned}$$

where the last equation follows from reindexing the terms $\tau(j)\sigma(j)$, $j = 1, \ldots, n$ in increasing order with respect to the values of $\tau(j)$. Of Course σ' is a permutation of I_n. This equation proves that the distribution of $\tilde{A}_\pi^n$ does not depend on the selection of the fixed permutation τ. Hence, from equation (25)

$$\hat{A}_\pi^n \overset{d}{=} \sum_{i=1}^n (i - \sigma(i))^2 = \sum_{i=1}^n 2(R_n - i\sigma(i)) = \sum_{i=1}^n Y_i, \qquad (26)$$

where $R_n = (n+1)(2n+1)/6$. So, we have expressed $\hat{A}_\pi^n$ as the sum of n random variables Y_i for $i = 1, \ldots, n$. Now, using Proposition 3.8, equation(15), for $i = 1, \ldots, n$ we obtain:

$$\begin{aligned}
\mathrm{E}(Y_i) &= 2R_n - 2i\mathrm{E}(\sigma(i)) \\
&= \frac{(n+1)(2n+1)}{3} - (n+1)i \\
&= (n+1)\left\{\frac{2n+1}{3} - i\right\}, \qquad (27)
\end{aligned}$$

$$\begin{aligned}
\mathrm{E}(Y_i^2) &= \mathrm{E}(4R_n^2 - 8R_n i\sigma(i) + 4i^2\sigma(i)^2) \\
&= 4\left\{\frac{(n+1)^2(2n+1)^2}{36} - \frac{i(n+1)^2(2n+1)}{6} + \frac{i^2(n+1)(2n+1)}{6}\right\} \\
&= \frac{2(n+1)(2n+1)}{3}\left\{\frac{(n+1)(2n+1)}{6} - i(n+1) + i^2\right\}. \qquad (28)
\end{aligned}$$

Therefore, using equations (27) and (28), for every $i \in \{1, 2, \ldots, n\}$

$$\begin{aligned}
\mathrm{Var}(Y_i) &= \mathrm{E}(Y_i^2) - (\mathrm{E}(Y_i))^2 \\
&= (n+1)i^2\left(\frac{2}{3}(2n+1) - (n+1)\right) \\
&= \frac{(n^2-1)i^2}{3}. \qquad (29)
\end{aligned}$$

By Proposition 3.8, equation (16), if $i \neq j$, $i, j \in \{1, 2, , \ldots, n\}$,

$$\begin{aligned}
\mathrm{E}\,(Y_iY_j) &= \mathrm{E}\,(4(R_n - i\sigma(i))(R_n - j\sigma(j))) \\
&= \mathrm{E}\,(4\{R_n^2 - R_n(i\sigma(i) + j\sigma(j)) + ij\sigma(i)\sigma(j)\}) \\
&= \frac{n+1}{3}\left\{\frac{(n+1)(2n+1)^2}{3} - (n+1)(2n+1)(i+j) + ij(3n+2)\right\} \qquad (30)
\end{aligned}$$

So, using equations (27) and (30),

$$\begin{aligned}
\mathrm{Cov}(Y_i, Y_j) &= \mathrm{E}\,(Y_iY_j) - \mathrm{E}\,(Y_i)\mathrm{E}\,(Y_j) \\
&= (n+1)ij\left(\frac{3n+2}{3} - (n+1)\right) \\
&= -\frac{(n+1)}{3}ij. \qquad (31)
\end{aligned}$$

Finally, using equations (29) and (31) we get

$$\begin{aligned}
\rho(Y_i, Y_j) &= \frac{\mathrm{Cov}(Y_i, Y_j)}{\sqrt{\mathrm{Var}(Y_i)\mathrm{Var}(Y_j)}} \\
&= -\frac{\frac{(n+1)ij}{3}}{\frac{(n+1)(n-1)ij}{3}} \\
&= -\frac{1}{(n-1)},
\end{aligned}$$

for every $i \neq j$, $i, j \in \{1, 2, \ldots, n\}$.

To prove the symmetry of $\hat{A}^n_\pi$, by equation (23) we know that $\hat{A}^n_\pi \stackrel{d}{=} \sum_{i=1}^n (i - \sigma(i))^2 = A$, but using equation (25)

$$A = \frac{n(n+1)(2n+1)}{3} - 2\sum_{i=1}^n i\sigma(i) \quad \text{for every} \quad \sigma \text{ permutation of } I_n.$$

For every σ fixed permutation of I_n, let $\tau(i) = n - \sigma(i) + 1$, for $i \in I_n$, which is obviously a permutation of I_n, and let $B = \sum_{i=1}^n (i - \tau(i))^2$. Then

$$\begin{aligned}
B &= \frac{n(n+1)(2n+1)}{3} - 2\sum_{i=1}^n (n - \sigma(i) + 1)i \\
&= \frac{n(n+1)(2n+1)}{3} - n^2(n+1) + 2\sum_{i=1}^n i\sigma(i) - n(n+1) \\
&= \frac{n(n+1)(2n+1)}{3} - n(n+1)^2 + 2\sum_{i=1}^n i\sigma(i).
\end{aligned}$$

Therefore,

$$\begin{aligned}
\frac{n^3-n}{3} - B &= \frac{n(n+1)(n-1)}{3} - \frac{n(n+1)(2n+1)}{3} + n(n+1)^2 - 2\sum_{i=1}^{n} i\sigma(i) \\
&= n(n+1)\left\{\frac{(n-1)}{3} - \frac{(2n+1)}{3} + (n=1)\right\} - 2\sum_{i=1}^{n} i\sigma(i) \\
&= \frac{n(n+1)(2n+1)}{3} - 2\sum_{i=1}^{n} i\sigma(i) \\
&= A.
\end{aligned}$$

Hence, $\hat{A}^n_\pi$ is symmetric with respect to $(n^3-n)/6$. □

The values of $\mathrm{E}\,(\hat{A}^n_\pi)$ and $\mathrm{Var}(\hat{A}^n_\pi)$ given in Proposition 3.8 can also be obtained using the results in Proposition 3.9.

From Proposition 3.9 we observe that the random variable $\hat{A}^n_\pi$ is a sum of n random variables Y_i, $i = 1, \ldots, n$, which are not identically distributed nor independent. However we also can observe that the random variables Y_i, $i = 1, \ldots, n$ are asymptotically uncorrelated. If we consider

$$A_n = \frac{\hat{A}^n_\pi - \mathrm{E}\,(\hat{A}^n_\pi)}{\sqrt{\mathrm{Var}(\hat{A}^n_\pi)}}. \tag{32}$$

Then A_n is a discrete symmetric random variable by Proposition 3.9, with $\mathrm{E}\,(A_n) = 0$ and $\mathrm{Var}(A_n) = 1$. Now using representation given in equation (26) we know that

$$\hat{A}^n_\pi \stackrel{d}{=} \sum_{i=1}^{n} (i - \sigma(i))^2.$$

If we let $Z_i = (i-\sigma(i))^2$ for $i = 1, \ldots, n$, then $\hat{A}^n_\pi \stackrel{d}{=} \sum_{i=1}^n Z_i$, and it can easily be proved that $Z_1 \stackrel{d}{=} Z_n$, $Z_2 \stackrel{d}{=} Z_{n-1}$, etc.

Now we will find the asymptotic distribution of A_n given in equation (32). Recall the sample version of Spearman's rho, see for example Nelsen [17] or Hettmansperger [5]. Let $\{(X_i, Y_i)\}_{i=1}^n$ be sample of a continuous joint distribution F. Let us assume that $X_1 < X_2 < \cdots < X_n$, and let

$$\rho = \frac{12}{n(n^2-1)}\left\{\sum_{i=1}^{n} iR_i - \frac{n(n+1)^2}{4}\right\}, \tag{33}$$

where $R_1, R_2, \ldots, R_n$ are the ranks of $Y_1, Y_2, \ldots, Y_n$. In our notation R_i corresponds to $\sigma(i)$ for $i = 1, 2, \ldots, n$. Now using equations (20), (17) and (18), we have that equation (32) becomes

$$\begin{aligned}
A_n &= \frac{\hat{A}^n_\pi - \mathrm{E}\,(\hat{A}^n_\pi)}{\sqrt{\mathrm{Var}(\hat{A}^n_\pi)}} \\
&= \frac{2\sum_{i=1}^n i\sigma(i) - \frac{n(n+1)(n+2)}{3} - \frac{n(n-1)(n+1)}{6}}{\sqrt{\frac{n^2(n+1)^2(n-1)}{36}}} \\
&= \frac{2\sum_{i=1}^n i\sigma(i) - \frac{n(n+1)}{6}[2n+4+n-1]}{\frac{n(n+1)}{6}\sqrt{n-1}} \\
&= \frac{12}{n(n+1)\sqrt{n-1}}\left(\sum_{i=1}^n i\sigma(i) - \frac{n(n+1)^2}{4}\right) \\
&= \sqrt{n-1}\cdot\frac{12}{n(n^2-1)}\left(\sum_{i=1}^n iR_i - \frac{n(n+1)^2}{4}\right) \\
&= \sqrt{n-1}\cdot\rho \qquad (34)
\end{aligned}$$

It is well known that under independence, $\sqrt{n-1}\rho$ is asymptotically $N(0,1)$, see Appendix, Section 5 of U-statistics in Lehman [13] or exercise 4.5.15 in Hettmansperger [5]. The proof of the asymptotical normality of $A_n = \sqrt{n-1}\rho$ in Lehmann [13] follows using U-statistics techniques and an approximation, on the other hand in Hettmansperger [5] the proof is based on the Projection Theorem and using Slutsky's Theorem. In both cases the proofs are elaborate.

As observed above the random variable A_n is just the standardization of the random variable $\hat{A}^n_\pi$, which is the sum of n random variables Y_i, $i = 1, \ldots, n$ standardized, which are not identically distributed nor independent according to Proposition 3.9, and they are asymptotically uncorrelated. However the sum of the random variables Y_i satisfy the central limit theorem. In fact, what is really surprising is the speed of convergence to a standard normal random variable. In Figure 4 we compare the exact distribution of A_n for $n = 12$ with the distribution of a standard normal variable, as can be seen for n as small as 12 the approximation to the standard normal is amazingly good.

Another way to observe the asymptotic normality of A_n would be to take n random permutations $\tau_1, \ldots, \tau_n$ and define $W_j = \sum_{i=1}^n (\tau_j(i) - \sigma(i))^2$, for $j = 1, 2, \ldots, n$. Then by Proposition 3.9 equation (23), we have n independent identically distributed random variables $W_1, \ldots, W_n$ and $W_1 \overset{d}{=} A_n$. Then using the usual central limit theorem we have that

$$T_n = \frac{1}{n}\sum_{j=1}^n\left(\sum_{i=1}^n (\tau_j(i) - \sigma(i))^2\right)$$

is asymptotically normal when standardized. Then we can compare the distributions of T_n and A_n.

The asymptotic distribution of A^n_π is also normal, this follows from Corollary 3.4, since for n large enough $\mathrm{P}(K = 1) \approx 1 - 2/n$, that is $A^n_\pi \approx \hat{A}^n_\pi$ for n large enough.

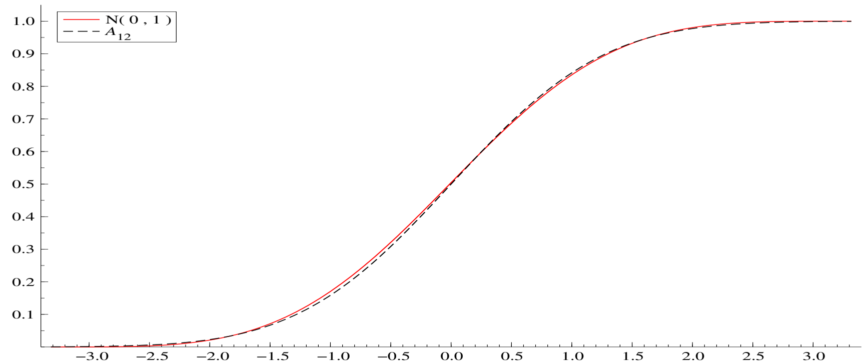

Fig. 4. Distribution functions of A_{12} and a standard normal.

Observe also that always $A_\pi^n \leq \hat{A}_\pi^n$, in Figure 5 we compare the exact densities of A_π^{12} and $\hat{A}_\pi^{12}$, as can be seen even for n small the densities are similar. In fact as n increases the density of A_π^n approaches "from the left" the density of $\hat{A}_\pi^n$.

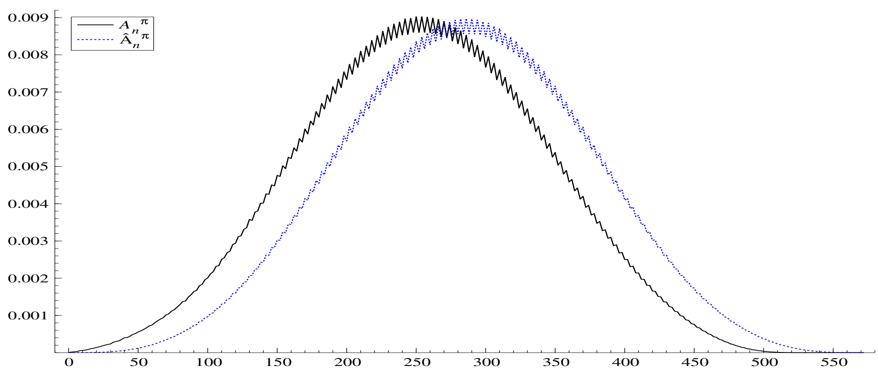

Fig. 5. Exact densities of A_π^{12} and $\hat{A}_\pi^{12}$.

4. ASSOCIATIVITY OF SAMPLES AND FINAL REMARKS

The distribution of the statistic A_π^n depends of course on the copula C from which we are sampling.

If we obtain a sample $(X_1, Y_1), \ldots (X_n, Y_n)$ from the product copula Π and we obtain its modified sample $(1, \sigma(1)), \ldots, (n, \sigma(n))$, we observe that any of the possible permutations σ of $\{1, 2, \ldots, n\}$ is equally likely to appear, since all permutations have the same probability $1/(n!)$ under independence.

We also notice that if the observed sample does not have idempotents, the statistic $A_\pi^n = \hat{A}_\pi^n$. Therefore, we are measuring the "l^2 distance" between the modified sample $\{(1,\sigma(1)),\ldots,(n,\sigma(n))\}$ and $\{(1,n),(2,n-1),\ldots,(n,1)\}$. Hence, if we are sampling from a copula with positive ρ correlation which is not too close to 1 (see next paragraph), we obtain large values of A_π^n. This is the case when we are sampling from Archimedean families which may vary between Π and M, such as the Clayton or Frank families with $\theta > 0$, see [17]. In [18], Nelsen studies copulas with maximal nonexchangeability or maximal asymmetry, this problem is also studied in [10], Nelsen proves that if X and Y are maximal nonexchangeable continuous random variables then $\rho_{X,Y} \in [-5/9, -1/3]$. This result opens an interesting question, that is, if maximal non associativity of two continuous random variables X and Y measured in some sense, leads to restrictions on the values of $\rho_{X,Y}$?

We also observe that if we are sampling from M, then our modified sample is $\{(1,1),(2,2),\ldots,(n,n)\}$ with probability one. Therefore, the sample has n idempotents and the value of $A_\pi^n = 0$ with probability one. In fact, the existence of some idempotents in the sample leads to small values of A_π^n, this is the case for copulas C which are very close to M.

In Lemma 3.7 we obtain the range of the statistic A_π^n, observe that the independence assumption is only used to find the probabilities of the minimum and maximum of A_π^n. In fact, Lemma 3.7 tells us which permutations σ lead to maximum values of the statistic A_π^n. This permutations allow us to construct copulas that generate samples with only large values of A_π^n, as we will see in the last example.

Let $\underline{X_n} = (X_1,Y_1),\ldots,(X_n,Y_n)$ be a random sample from an Archimedean copula C, and let $(1,\sigma(1)),\ldots,(n,\sigma(n))$ be the modified sample defined in Section 2. We perform a simulation study of the behavior of A_π^n under different families, see [17], such as the Clayton, Frank, Gumbel-Barnett, etc., and we compare the simulated distributions to the distribution obtained sampling from Π. For example, in the case of the Clayton family C_θ, with $\theta \in [-1,\infty)\backslash\{0\}$, which includes as limiting cases $C_{-1} = W$, $C_0 = \Pi$ and $C_\infty = M$, and as a special case $C_1 = \Pi/(\Sigma - \Pi)$. We observed that the value of Spearman's rho increases from -1 to 1 when θ varies from -1 to ∞, and tends to 0 when θ approaches 0. We generated 10000 samples of size $n = 100$ of the Clayton family varying θ from -0.9 up to 50 obtaining the empirical distributions of A_π^{100} in each case. In Figure 6 we include the empirical distributions of A_π^{100} for $\theta = -0.7, -0.3, 0, 0.5, 1, 3, 8$ and 30. As can be seen from this graph the distribution of A_π^n moves to the right when θ increases from -0.7 up to 3, for the last two values of the parameter θ the distribution moves to the left. We also observe that for $\theta = 3$, the value of Spearman's rho equals $\rho = .78$ and in this case the distribution still moves to the right, but its variance starts to grow. For $\theta = 8$ Spearman's rho equals $\rho = .94$, and in this case the distribution moves to the left and has a larger variance. Finally, for $\theta = 30$ the distribution of A_π^{100} approaches the distribution of the constant zero. Of course for the limiting cases M and W the distribution is degenerate into the constant zero. A similar situation is obtained if we are sampling from the Frank family.

Finally, we would like to construct copulas which produce very little associative samples.

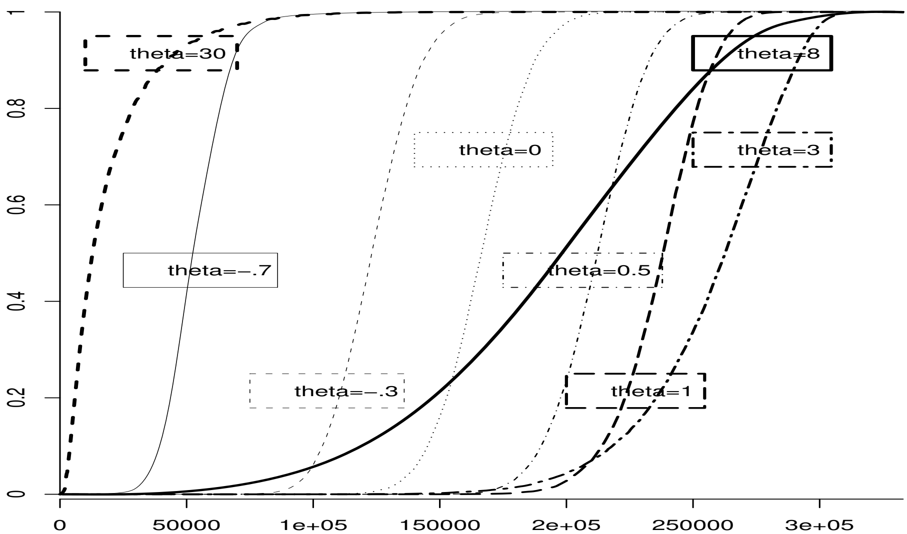

Fig. 6. Distribution functions of A_π^{100} for the Clayton family.

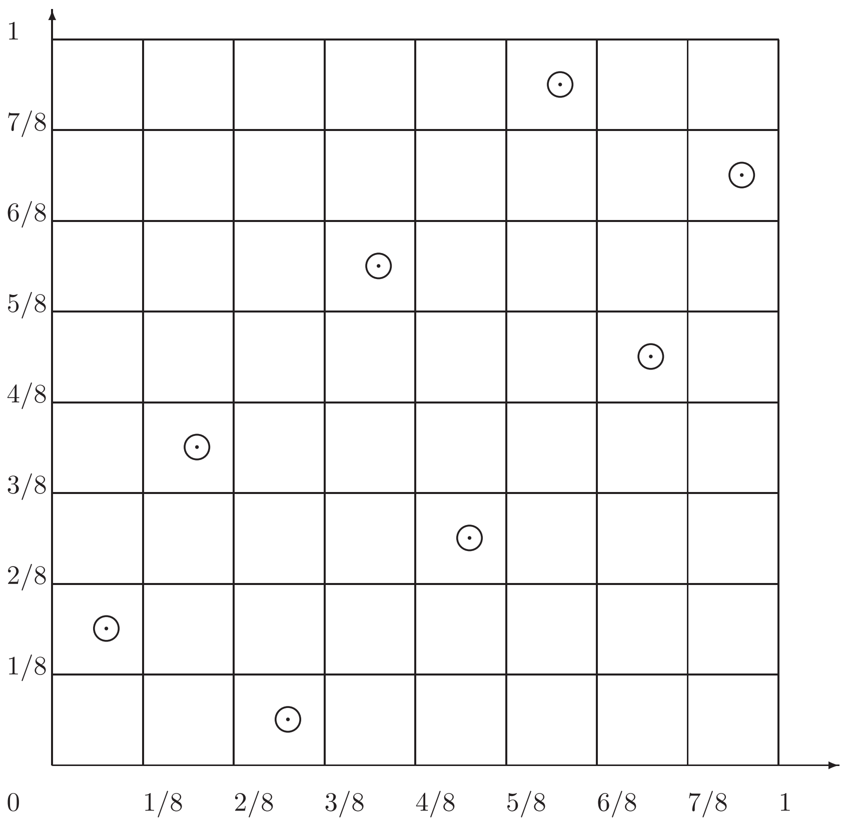

Fig. 7. Non-symmetric non-assoc. copula with mass on dotted squares.

Recall that in the proof of Lemma 3.7, where we defined the range of the statistic A_π^n, we found for each n, the least associative samples according to the maximum value of A_π^n. If we let the sample size to be 8, we observed that the sample $\{(1,2),(2,4),(3,1),(4,6),\ (5,3),(6,8),(7,5),(8,7)\}$ is one of the two least associative samples. As an example, and based on the observations above, we generated 100000 simulations of sample sizes $n = 25$, $n = 50$ and $n = 100$ of a copula $C(u,v)$ whose support is uniform on the little squares given in Figure 7, that is, the density $c(u,v) = \frac{\partial^2}{\partial u \partial v} C(u,v)$ associated to this copula is given by $c(u,v) = 8$, if (u,v) belongs to any of the little dotted squares in Figure 7, and $c(u,v) = 0$ otherwise. For the three sample sizes we obtained the empirical distributions, we observe that as n increases the values of A_π^n concentrate more and more on large values that are close to the maximum value of A_π^n.

The other least associative sample of size 8 is given by $\{(1,3),(2,1),(3,5),(4,2),$ $(5,7),(6,4),(7,8),(8,6)\}$, which corresponds to the symmetric version of the other sample. We also performed the simulations obtaining in this case similar results.

- This paper introduces a new statistic A_π^n which is nonnegative and measures the associativity of given sample using a characterization of associativity for discrete copulas. The larger the value of the statistic the less associative the sample.

- Given any value of the sample size n and using Lemma 3.7, we can say exactly which modified samples produce the largest possible value of the statistic A_π^n, hence the less associative samples.

- Since the composite problem of associativity is quite complex, and the statistic A_π^n is not distribution free, using an appropriate representative member of the associative copulas, that is Π, it is possible to find its exact distribution for small sample sizes n when the number of idempotents of the sample is greater than one, see Theorem 3.6.

- In the case that the sample has only the trivial idempotent $i_1 = n$, the statistic A_π^n reduces to the statistic $\hat{A}_\pi^n$, which under standardization using Proposition 3.8, reduces to the Spearman's ρ. Observe that at least in this case, we find a nice statistical interpretation of associativity of samples, answering at least partially the question raised by Nelsen [17], see Section 2.

- In the case of independence it is also possible to find the normal asymptotic distribution not only for the simplified statistic $\hat{A}_\pi^n$, but also for the original statistic A_π^n, using Corollary 3.4 and Table 1.

ACKNOWLEDGEMENT

I would like to thank Arturo Erdely for several interesting conversations about this problem. This work was partially supported by Mexico's CONACyT grant 50152. I would like to thank two anonymous referees for a careful revision of the first manuscript, and their interesting references and statistical observations which improved this paper.

(Received December 11, 2008)

REFERENCES

[1] I. Aguiló, J. Suñer, and J. Torrens: Matrix representation of discrete quasi-copulas. Fuzzy Sets and Systems (2007), doi: 10.1016/j.fss2007.10.004.

[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] A. Erdely, J. M. González-Barrios, and R. B. Nelsen: Symmetries of random discrete copulas. Kybernetika *44* (2008), 6, 846–863.

[5] T. P. Hettmansperger: Statistical Inference Based on Ranks. Wiley, New York 1984.

[6] S. Jenei: On the convex combination of left-continuous t-norms. Aequationes Mathematicae *72* (2006), 47–59.

[7] S. Jenei: On the geometry of associativity. Semigroup Forum *74* (2007), 439–466.

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

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

[10] E. P. Klement and R. Mesiar: How non-symmetric can a copula be? Comment. Math. Univ. Carolinae *47* (2006), 1, 141–148.

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

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

[13] E. L. Lehmann: Nonparametric Statistical Methods Based on Ranks. Revised first edition. Springer, New York 2006.

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

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

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

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

[18] R. B. Nelsen: Extremes of nonexchangeability. Statist. Papers *48* (2007), 329–336.

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

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.
e-mail: gonzaba@sigma.iimas.unam.mx