Arturo Erdely; José M. González–Barrios

Exact distribution under independence of the diagonal section of the empirical copula

Kybernetika, Vol. 44 (2008), No. 6, 826--845

Persistent URL: http://dml.cz/dmlcz/135894

KYBERNETIKA — VOLUME 44 (2008), NUMBER 6, PAGES 826 – 845

EXACT DISTRIBUTION UNDER INDEPENDENCE OF THE DIAGONAL SECTION OF THE EMPIRICAL COPULA

Arturo Erdely and José M. González–Barrios

In this paper we analyze some properties of the empirical diagonal and we obtain its exact distribution under independence for the two and three-dimensional cases, but the ideas proposed in this paper can be carried out to higher dimensions. The results obtained are useful in designing a nonparametric test for independence, and therefore giving solution to an open problem proposed by Alsina, Frank and Schweizer [2].

Keywords: Archimedean copula, diagonal section, independence

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

1.INTRODUCTION

A copula C is said to be *Archimedean*, see Nelsen [24], if $C(u,v) = \varphi^{[-1]}[\,\varphi(u) + \varphi(v)\,]$, where φ is called the *generator* of the copula, which is a continuous, convex, strictly decreasing function from $[\,0,1\,]$ to $[\,0,\infty\,]$ such that $\varphi(1) = 0$, and $\varphi^{[-1]}$ is the *pseudo-inverse* of φ. Its *diagonal section* is given by

$$\delta_C(u) = C(u,u) = \varphi^{[-1]}[\,2\varphi(u)\,]\,. \tag{1}$$

Therefore if we know the form of the generator φ then it is straightforward to obtain δ_C. One may ask, as observed by Darsow and Frank [8], how much information about an Archimedean copula is contained in its diagonal section. In other words, given δ, what can be said about φ? We may rewrite (1) as

$$\varphi\big(\delta(u)\big) = 2\,\varphi(u)\,, \qquad u \in [\,0,1\,]\,. \tag{2}$$

The above equation is a particular case of *Schröder's functional equation,* which has been studied in one form or another, according to Kuczma [19], since the late nineteenth century. Frank [13] announced that a sufficient, but not necessary condition for an Archimedean copula C to be uniquely determined by its diagonal section δ is given by the left derivative $\delta\,'(1-) = 2$, which we will call *Frank's condition.* This is an almost immediate consequence of standard results on convex solutions of Schröder's equation via the representation of these copulas. Frank also illustrated that this condition is not necessary by constructing families of Archimedean copulas having identical diagonals. The problem of finding conditions under which an Archimedean copula is or is not uniquely determined by its diagonal section has been also studied in the context of *triangular norms*, see for example Klement, Mesiar and Pap [15] and Klement and Mesiar [16].

Whenever Frank's condition is satisfied, we may apply Kuczma's [19] Theorem 6.6 to obtain the following formula for φ in terms of diagonal δ :

$$\varphi(u) = \lim_{n \to \infty} 2^{\,n}\big[1 - \delta^{\,-n}(u)\big]\,,$$

where $\delta^{\,-n}$ is the composition of $\delta^{\,-1}$ with itself n times. An important example of an Archimedean copula that satisfies Frank's condition is the case of the product copula $\Pi(u,v) = uv$, which characterizes a couple of independent continuous random variables, via Sklar's Theorem [26], and so it is uniquely determined by its diagonal section $\delta_{\,\Pi}(u) = u^2$.

As a consequence of the above results, Alsina, Frank and Schweizer [2] included in their book as an open problem, the following:

> Can one design a test of statistical independence based on the assumptions that the copula in question is Archimedean and that its diagonal section is $\delta(u) = u^2$?

For this purpose, first we recall Sklar's Theorem [26] and an immediate corollary:

Theorem. (Sklar) Let X and Y be random variables with distribution functions F and G , respectively, and joint distribution function H . Then there exists a copula C such that

$$H(x,y) = C(F(x), G(y))\,.$$

If F and G are continuous, C is unique; otherwise, C is uniquely determined on $Ran\,F \times Ran\,G$.

Corollary. Let X and Y be continuous random variables. Then X and Y are independent if and only if their corresponding copula is $C(u,v) = uv$.

It is customary to use the notation $\Pi(u,v) := uv$ and to call it the *product* or *independence copula*. The previous results imply that the product copula characterizes independent random variables when the distribution functions are continuous. The product copula is Archimedean and it is characterized by the diagonal section $\delta_{\,\Pi}(u) = u^2$, since it satisfies Frank's condition.

An answer to the question above implies to study the probability distribution of the empirical diagonal, under the (null) hypothesis of interest, that is, under independence. In the present work, we prove that it is possible to obtain the *exact probability distribution* of the empirical diagonal, under the hypothesis of independence. This opens the door to define suitable test statistics based on the empirical diagonal, for a nonparametric test for independence, with the advantage that their exact distribution will be also known, and therefore such a test will be also useful under small samples, in contrast with other nonparametric tests which rely on the asymptotic behavior of their proposed statistics.

2.THE EMPIRICAL DIAGONAL AND SOME PROPERTIES: BIVARIATE CASE

We have seen so far that, in the case of Archimedean bivariate copulas, the diagonal section contains all the information we need to build the copula in case Frank's condition $\delta'(1-) = 2$ is fulfilled, and in such case this leads us to concentrate in studying and estimating the diagonal. The main benefit of this fact is a reduction in the dimension of the estimation, from 2 to 1 in the case of bivariate copulas, and from m to 1 in the case of m-variate copulas.

Let $S := \{(X_1, Y_1), \ldots, (X_n, Y_n)\}$ denote a random sample of size n from a continuous random vector (X, Y) . As defined by Deheuvels [9], the *(bivariate) empirical copula* is the function C_n given by

$$C_n\left(\frac{i}{n}, \frac{j}{n}\right) = \frac{1}{n} \sum_{(X\,Y) \in S} \mathbf{1}_{]-\infty\,,\,X_{(i)}\,]\times]-\infty\,,\,Y_{(j)}\,]}(X, Y)\,,$$

where $X_{(i)}$ and $Y_{(j)}$ denote the order statistics of the sample, for i and j in $\{1, \ldots, n\}$, and $C_n(\frac{i}{n}, 0) = 0 = C_n(0, \frac{j}{n})$. The domain of the empirical copula is the grid $\{0, 1/n, \ldots (n-1)/n, 1\}^2$ and its range is the set of $\{0, 1/n, \ldots, (n-1)/n, 1\}$. The domain of the empirical copula is just a rescaling of the set $\{0, 1, \ldots, n\}$. Hence the empirical copula can be thought as equivalent to a discrete copula as noticed in Mayor et al. [21] and Mesiar [23]. Moreover, an empirical copula is an example of an irreducible discrete copula as defined in Kolesárová et al. [17]. An empirical copula is not a copula, but a (two-dimensional) *subcopula*, for details of subcopulas see Nelsen [24].

Definition 2.1. The bivariate empirical diagonal is the function δ_n given by

$$\delta_n\left(\frac{j}{n}\right) := C_n\left(\frac{j}{n}, \frac{j}{n}\right) \quad j = 0, 1, \ldots, n\,.$$

Without loss of generality we may assume that the X_k values in S are ordered, then

$$\delta_n\left(\frac{j}{n}\right) = \frac{1}{n} \sum_{k=1}^{j} \mathbf{1}_{]-\infty\,,\,Y_{(j)}\,]}(Y_k),\,, \quad j = 1, \ldots, n-1\,, \tag{3}$$

and $\delta_n(0) = 0$, $\delta_n(1) = 1$. It is clear from above that δ_n is a nondecreasing function of j . Moreover, by the Fréchet–Hoeffding bounds for subcopulas:

$$\max\left(\frac{2j}{n} - 1, 0\right) \le \delta_n\left(\frac{j}{n}\right) \le \frac{j}{n}\,. \tag{4}$$

We can also prove that

$$\delta_n\left(\frac{j+1}{n}\right) - \delta_n\left(\frac{j}{n}\right) \in \left\{0, \frac{1}{n}, \frac{2}{n}\right\},$$

because

$$\begin{aligned}
\delta_n\left(\frac{j+1}{n}\right) - \delta_n\left(\frac{j}{n}\right) &= \frac{1}{n}\Bigg[\mathbf{1}_{]-\infty\,,\,Y_{(j+1)}\,]}(Y_{j+1}) \\
&\quad + \sum_{k=1}^{j}\left(\mathbf{1}_{]-\infty\,,\,Y_{(j+1)}\,]}(Y_k) - \mathbf{1}_{]-\infty\,,\,Y_{(j)}\,]}(Y_k)\right)\Bigg], \\
&= \frac{1}{n}\left[\mathbf{1}_{]-\infty\,,\,Y_{(j+1)}\,]}(Y_{j+1}) + \sum_{k=1}^{j}\mathbf{1}_{]\,Y_{(j)}\,,\,Y_{(j+1)}\,]}(Y_k)\right], \\
&= \frac{1}{n}\left[\mathbf{1}_{]-\infty\,,\,Y_{(j+1)}\,]}(Y_{j+1}) + \sum_{k=1}^{j}\mathbf{1}_{\{Y_{(j+1)}\}}(Y_k)\right].
\end{aligned}$$

So this means that all the possible paths $\{\delta_n(\frac{j}{n}) : j = 0, 1, \ldots, n\}$ are between the paths $\{\max(\frac{2j}{n} - 1, 0) : j = 0, 1, \ldots, n\}$ and $\{\frac{j}{n} : j = 0, 1, \ldots, n\}$ with jumps of size 0, $\frac{1}{n}$, or $\frac{2}{n}$ between consecutive steps. This also follows from properties of the diagonal section in discrete copulas and quasi-copulas, see Aguiló et al. [1] or Kolesárová and Mordelová [18].

Let X be a continuous uniform random variable in $]\,0,1\,[$ and define the random variable $Y := X$. Then the corresponding copula for (X, Y) is the Fréchet–Hoeffding upper bound copula $M(u, v) := \min(u, v)$. In this case, a size n sample of observations of (X, Y) would be $S = \{(X_1, X_1), \ldots, (X_n, X_n)\}$, and applying formula (3) we get $\delta_n(\frac{j}{n}) = \frac{j}{n}$, which is the Fréchet–Hoeffding upper bound in (4). If, instead, we define $Y := 1 - X$, the corresponding copula for (X, Y) is the Fréchet–Hoeffding lower bound copula $W(u, v) := \max(u + v - 1, 0)$, and $\delta_n(\frac{j}{n})$ equals the lower bound in (4).

Definition 2.2. Let $(X_1, Y_1), \ldots, (X_n, Y_n)$ be a random sample of the random vector (X, Y) of continuous random variables, where $X_1 < X_2 < \cdots < X_n$. Let us define the diagonal random path by the vector

$$\boldsymbol{T} = \left(\delta_n\left(\frac{0}{n}\right), \delta_n\left(\frac{1}{n}\right), \ldots, \delta_n\left(\frac{n}{n}\right)\right)$$

where

$$\delta_n\left(\frac{j}{n}\right) := \frac{1}{n}\sum_{k=1}^{j}\mathbf{1}_{]-\infty\,,\,Y_{(j)}\,]}(Y_k)\,, \quad j = 1, \ldots, n-1\,,$$

with $\delta_n(0) = 0$, $\delta_n(1) = 1$.

Associated to the diagonal random path we may define the *diagonal random increments* by the vector $\boldsymbol{I} := \left(\delta_n\left(\frac{1}{n}\right) - \delta_n\left(\frac{0}{n}\right), \delta_n\left(\frac{2}{n}\right) - \delta_n\left(\frac{1}{n}\right), \ldots, \delta_n\left(\frac{n}{n}\right) - \delta_n\left(\frac{n-1}{n}\right)\right)$ so that knowledge of $\boldsymbol{T} = (t_0 = 0, t_1, \ldots, t_{n-1}, t_n = 1)$ is equivalent to knowledge of $\boldsymbol{I} = (i_1, \ldots, i_n)$, where $i_j = t_j - t_{j-1}$ and $t_j = \sum_{k=1}^{j} i_k$.

Alternatively, we may write $\boldsymbol{I} = (i_1, \ldots, i_n) = \frac{1}{n}(b_1, \ldots, b_n)$ where the $b_j \in \{0, 1, 2\}$. Moreover, $i_n = 1 - t_{n-1} = 1 - \sum_{k=1}^{n-1} i_k$ so knowledge of $\boldsymbol{B} := (b_1, \ldots, b_{n-1})$ completely specifies any path. Different values of $\boldsymbol{B}$ can be labeled as vectors of ternary numbers. For example, with $n = 7$ the Fréchet–Hoeffding lower bound path $\{\max(\frac{2j}{n} - 1, 0) : j = 0, 1, \ldots, n\}$ is specified by the vector $(0, 0, 0, 1, 2, 2, 2)$, while the Fréchet–Hoeffding upper bound path $\{\frac{j}{n} : j = 0, 1, \ldots, n\}$ is specified by the vector $(1, 1, 1, 1, 1, 1, 1)$.

We will call an *admissible diagonal path* any vector of ternary numbers satisfying the Fréchet–Hoeffding conditions. We will now find the exact number of admissible paths for any $n \geq 2$. For this purpose, we have to recall the problem of walks on the integral lattice, where we look at walks of $m + k$ unit steps into upward and rightward directions, starting at the origin $(0, 0)$ and ending at (m, k). The number of such paths without further restrictions is $\binom{m+k}{k}$, as exactly k of the $m + k$ steps are upward steps. Now consider just those upward-rightward paths with $k \leq m$, that is, paths remaining on or under the diagonal. For this to happen it is necessary to have at any step of the path an accumulated number of rightward steps equal or larger than the number of upward steps: "a rightward step before any upward step."

For the case $k = m$ it is proved in Theorem 3.1 in Barcucci and Verri [4] that $E_m := \frac{1}{m+1}\binom{2m}{m}$ is the number of the under-diagonal rightward-upward one-step walks on the integral lattice (the sequence $\{E_m\}$ is known as the *Catalan numbers*). An equivalent result is the classical *Chung–Feller Theorem*, see Chung and Feller [7], or Feller [12] in his chapter *On fluctuations in coin-tossing and random walks.* For the general case, by a result in Bailey [3], we may calculate the number of under-diagonal rightward-upward one-step walks on the integral lattice, starting at $(0, 0)$ and ending at (m, k), by

$$\frac{m+1-k}{m+1}\binom{m+k}{m}, \quad k \leq m\,. \tag{5}$$

In his original version, Bailey [3] obtains (5) by counting the number of sequences with non-negative partial sums that consist of m positive 1's and k negative 1's. An equivalent result is found in Engleberg [10].

Proposition 2.3. Let P_n denote the number of admissible paths for the bivariate empirical diagonal δ_n of n points in $[0, 1]^2$ and $n \geq 2$. Then

$$P_n = \sum_{r_0=0}^{[\![n/2]\!]} \frac{\binom{n}{r_0}\binom{n-r_0}{r_0}}{r_0 + 1}$$

where $[\![\, x \,]\!]$ denotes the greatest integer less than or equal to x.

P r o o f. Let $(x_1, y_1), \ldots, (x_n, y_n) \in [0,1]^2$ be n points and let δ_n be their empirical diagonal. Let r_0, r_1 and r_2 denote the number of zeros, ones and twos respectively in the empirical diagonal $n\delta_n$. Then $r_0 + r_1 + r_2 = n$ and $r_1 + 2r_2 = n$, so $r_0 = r_2$ by (18). Hence

$$r_1 + 2r_0 = n, \tag{6}$$

any nonnegative integers r_0 and r_1 satisfying (6), could provide an admissible path for the empirical diagonal whenever the Fréchet–Hoeffding bounds are satisfied. Observe that $r_0 \leq [\![n/2]\!]$.

If $\delta_n = \frac{1}{n}(b_1, b_2, \ldots, b_n)$ is an admissible path, then $b_1 = 0$ or $b_1 = 1$, and for $2 \leq i \leq n$ $b_i = 0, 1$ or 2. We observe that the restrictions $\sum_{i=1}^{n} b_i = n$ and $\sum_{i=1}^{j} b_i \leq j$ must be fulfilled. The *basic rule* to find admissible paths is "zero before two." That is if some $b_i = 2$, then there exists $1 \leq j < i$ such that $b_j = 0$. For example if $n = 5$ and $r_0 = 1$, then $(0,1,1,2,1)$ is an admissible path, but $(2,1,1,0,1)$ is not admissible, since for example $\sum_{i=1}^{3} b_i = 4 \not\leq 3$. Therefore, given the number of zeros among the b_i's, we only have to see where they can be located in the vector $(b_1, \ldots, b_n)$, following the basic rule. Observe that the ones can be located in any place.

So first assume that $r_0 = 0$ and $r_1 = n$, then the only path, which by the way is admissible, is $(1, 1, \ldots, 1)$, that is, Fréchet–Hoeffding upper bound path. Now fix some r_0 such that $1 \leq r_0 \leq [\![n/2]\!]$. We have to count all the admissible paths that follow the basic rule "zero before two." Since the ones can be located any place, we just have to count the different ways in which we can locate the zeros and the twos. First, we have to choose $r_0 + r_2 = 2r_0$ places for the zeros an twos out of the n places available, which can be made in $\binom{n}{2r_0}$ different ways. This last number has to be multiplied by the number of different ways in which we can locate the r_0 zeros and the $r_2 = r_0$ twos in the $2r_0$ chosen places, but always following the basic rule. We may relate the zeros to rightward unit steps and the twos to upward unit steps in (5) with $m = k = r_0$, so the number of admissible paths with r_0 zeros is given by

$$\binom{n}{2r_0}\binom{2r_0}{r_0}\frac{1}{r_0+1} = \frac{1}{r_0+1}\binom{n}{r_0}\binom{n-r_0}{r_0}. \tag{7}$$

The result now follows summing over all possible values of r_0. $\square$

Remark. (7) also simplifies in terms of a multinomial coefficient to

$$\frac{1}{r_0+1}\binom{n}{r_0,\, r_1,\, r_2}, \tag{8}$$

where $r_2 = r_0$ and $r_0 + r_1 + r_2 = n$. This may be understood as follows: the multinomial coefficient along with the restriction $r_0 + r_1 + r_2 = n$ represents the number of permutations of repeated elements (r_0 zeros, r_1 ones, and r_2 twos, in this case), but some of these permutations do not follow the basic rule "zero before two", so the role of the factor $\frac{1}{r_0+1}$ is to leave just those permutations that follow the rule. To understand how this factor is obtained, we need some combinatorics concepts as in, for example, Callan [5], considering the problem of building sequences of 1's and -1's, and their partial sums: A *balanced n-path* is a sequence of n [1's] and n [-1's], represented as a path of unit upsteps $(1,1)$ and downsteps $(1,-1)$ from $(0,0)$ to $(2n,0)$. A *Dick n-path* is a balanced n-path that never drops below the x-axis (ground level) "·" the parameter X on balanced n-paths defined by $X =$ "number of upsteps above ground level" is uniformly distributed over $\{0,1,2,\ldots,n\}$ and hence divides the $\binom{2n}{n}$ balanced n-paths into $n+1$ equal-size classes, one of which consists of the Dick n-paths (the one with $X = n$). Indeed, for $1 \le i \le n$, a bijection from balanced n-paths with $X = 0$ (inverted Dick paths) to those with $X = i$ is as follows. Number the upsteps from right to left and top to bottom, starting with the last upstep. Then remove the first downstep $[-1]$ encountered directly west of upstep i to obtain the subpaths P and Q, and reassemble as $Q\,[-1]\,P$.

So the problem of counting the different ways of allocating the zeros and twos following the basic rule is the same as counting the number of Dick n-paths, using the 1's to represent zeros, and the -1's to represent twos (remember that the ones may be allocated any place, so they will be allocated in the remaining places), and the result (8) follows.

Now we will calculate the probability of any given (admissible) path, under the hypothesis of independence.

Theorem 2.4. Let $S = \{(X_1,Y_1),\ldots,(X_n,Y_n)\}$ be a random sample from the random vector of continuous random variables (X,Y). If X and Y are independent and if $\boldsymbol{T} = (t_0 = 0, t_1, \ldots, t_{n-1}, t_n = 1)$ is an admissible diagonal path. Then

$$\Pr\left[\,\boldsymbol{T} = (t_0 = 0, t_1, \ldots, t_{n-1}, t_n = 1)\,\right] = \frac{1}{n!}\prod_{j=1}^{n} f(j)\,,$$

where $f(j)$ is obtained in terms of the following formula, for $j = 1,\ldots,n$:

$$f(j) = \begin{cases} 1 & \text{if} \quad n(t_j - t_{j-1}) = 0\,, \\ 2(j - n\,t_{j-1}) - 1 & \text{if} \quad n(t_j - t_{j-1}) = 1\,, \\ (j - 1 - n\,t_{j-1})^2 & \text{if} \quad n(t_j - t_{j-1}) = 2\,. \end{cases}$$

Proof. From the continuity assumption we know that with probability one there are no ties among the X_i's or the Y_i's. Without loss of generality we may assume that the X_k $(k = 1,\ldots,n)$ are ordered, so we have that, by independence of X and Y, the probability of the random sample $\{(X_1,Y_1),\ldots,(X_n,Y_n))\}$ equals that of $\{(X_1,Y_{\sigma(1)}),\ldots,(X_n,Y_{\sigma(n)})\}$ where $\sigma(1),\ldots,\sigma(n)$ is any permutation of $(1,\ldots,n)$, and every permutation has probability $(n!)^{-1}$.

By rescaling we can assume that $X_i = i$, for $i = 1,2,\ldots,n$ and $Y_{\sigma(i)} = \sigma(i)$, for $i = 1,\ldots,n$. Hence the sample S is a subset of the grid $\{1,2,\ldots,n\}\times\{1,2,\ldots,n\} := I_n^2$. In fact, for every $i \in \{1,2,\ldots,n\}$ there exists a unique $j = \sigma(i) \in \{1,2,\ldots,n\}$ such that $(i,\sigma(i)) \in S$. That is for any horizontal or vertical segment in the grid I_n^2 there is exactly one point that belongs to the sample S.

In order to calculate $\Pr\left[\boldsymbol{T} = (0, t_1, \ldots, t_{n-1}, 1)\right]$ we just need to count the number of orderings of $\{Y_1, \ldots, Y_n\}$ that would lead to the admissible path $(t_0 = 0, t_1, \ldots, t_{n-1}, t_n = 1)$. We will show that this probability is given by $\frac{\prod_{j=1}^{n} f(j)}{n!}$.

Let $\boldsymbol{B} = (b_1, b_2, \ldots, b_n)$ be the vector of ternary numbers which is equivalent to the admissible diagonal path $\boldsymbol{T} = (t_0 = 0, t_1, \ldots, t_{n-1}, t_n = 1)$, that is $b_i = n(t_i - t_{i-1})$ for $i = 1, 2, \ldots, n$. Define

$$K = \min\{i \in \{1, \ldots, n\} \mid b_i > 0\}.$$

Since $\boldsymbol{B}$ is an admissible diagonal path, with all the $b_i's$ being equal to $0, 1$ or 2, except for b_1 which equals 0 or 1, and $\sum_{i=1}^{n} b_i = n$. Then $K \leq [\![n/2]\!]$. Therefore if $K = 1$ it means that $b_1 = 1$, and then $(1, 1) \in S$, and there is only one possibility for $\sigma(1)$, that is $\sigma(1) = 1$.

So, assume that $K > 1$, from the definition of the empirical copula it is clear that

$$nt_i = nC_n\left(\frac{i}{n}, \frac{i}{n}\right) = \mathrm{card}(S \cap (\{1, \ldots, i\} \times \{1, \ldots, i\})) \quad \text{for} \quad i = 1, 2, \ldots, n,$$

where card$(\cdot)$ stands for cardinality of a set. So nt_i gives us the number of sample points in the sample S that belong to the sub-grid $\{1, \ldots, i\}^2$. Since $K > 1$, then $b_1 = \cdots = b_{K-1} = 0$, which is equivalent to $t_1 = \cdots = t_{K-1} = 0$.

Now, first assume that $b_K = 1$, or equivalently $t_k = 1$. Then we observe that the intersection of the sub-grid $\{1, \ldots, K-1\} \times \{1, \ldots, K-1\}$ and the sample S is empty, but the intersection of the sub-grid $\{1, \ldots, K\} \times \{1, \ldots, K\}$ and the sample S contains a unique point. By noticing that $(\{1, \ldots, K\} \times \{1, \ldots, K\}) \backslash (\{1, \ldots, K-1\} \times \{1, \ldots, K-1\}) = \{(1, K), (2, K), \ldots, (K, K), (K, K-1), \ldots, (K, 1)\}$ we can select the point of the sample in $2K - 1 = 2(K - nt_{K-1}) - 1$ forms. If for example we select the point $(2, K)$ then we know that any point of the form $(2, j)$ for $j \neq K$, and any point of the form (l, K) with $l \neq 2$ do not belong to the sample, that is, we cancel one column and one row and the remaining points that were not selected.

The other possibility is $b_K = 2$. Then we observe that the intersection of the sub-grid $\{1, \ldots, K-1\} \times \{1, \ldots, K-1\}$ and the sample S is empty, but the intersection of the sub-grid $\{1, \ldots, K\} \times \{1, \ldots, K\}$ and the sample S contains exactly two points. Just as above we know that, $(\{1, \ldots, K\} \times \{1, \ldots, K\}) \backslash (\{1, \ldots, K-1\} \times \{1, \ldots, K-1\}) = \{(1, K), (2, K), \ldots, (K, K), (K, K-1), \ldots, (K, 1)\}$ contains two points of the sample S. Observe that (K, K) can not be a sample point, since in that case, none of the points $(1, K), \ldots (K-1, K), (K, K-1), \ldots, (K, 1)$ can belong to the sample. Therefore we can select one point from $(1, K), \ldots, (K-1, K)$ and another from $(K, K-1), \ldots, (K, 1)$, that is we have $(K-1)^2 = (K - 1 - nt_{K-1})^2$ possible choices. After selecting these two points we can not repeat the same indexes for columns or rows, so we cancel two columns and two rows and the remaining points which were not selected. Now we define

$$K_1 = \min\{i \in \{K+1, \ldots, n\} \mid b_i > 0\},$$

and we proceed inductively by reducing the dimension of the grid. As an example consider that $n = 5$, and the admissible path is given by $\boldsymbol{T} = (0 = t_0, t_1 = 0,$

$t_2 = 0, t_3 = 1/5, t_4 = 3/5, t_5 = 5/5 = 1)$, or equivalently $\boldsymbol{B} = (b_1 = 0, b_2 = 0, b_3 = 1, b_4 = 2, b_5 = 2)$, in this case

$$K = \min\{i \in \{1, \ldots, 5\} \,|\, b_i > 0\} = 3.$$

We first notice that $(1,1), (1,2), (2,1)$ and $(2,2)$ are not sample points, since $K = 3$, see Figure 1.

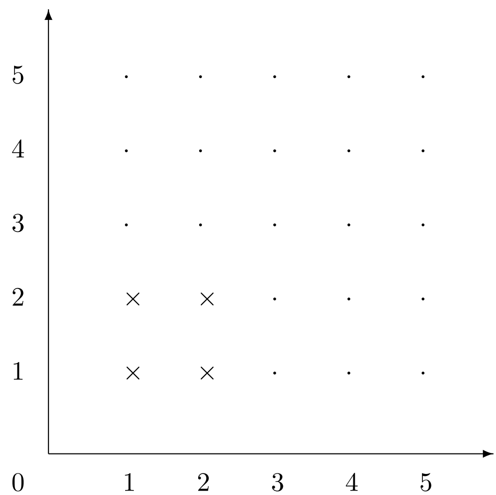

Fig. 1.

Now, $b_3 = 1$, so we have to select only one point in the set $\{(1,3), (2,3), (3,3), (3,2), (3,1)\}$, that is we have $5 = 2(3 - 5t_2) - 1$ choices. Assume we select $(1,3)$, then we cancel the remaining elements not selected and those of the first column and the third row, see Figure 2.

Now

$$K_1 = \min\{i \in \{4,5\} \,|\, b_i > 0\} = 4,$$

and $b_4 = 2$, in this case we have to select one point between $(2,4)$ and $(3,4)$ and another between $(4,1)$ and $(4,2)$, that is $2^2 = (4 - 1 - 5t_3)^2 = (3 - 5(1/5))^2$ choices, recall that $(4,4)$ can not be selected. Assume we select $(3,4)$ and $(4,1)$, so we cancel the third and fourth columns and the first and fourth row, and the points that were not selected, see Figure 3.

Finally,

$$K_2 = \min\{i \in \{5\} \,|\, b_i > 0\} = 5,$$

since $b_5 = 2$ we have two select two points between $(2,5)$ and $(5,2)$ recall that $(5,5)$ cannot be selected, this can be done only in $1 = 1^2 = (5-1-5t_4)^2 = (5-1-5(3/5))^2$ way, see Figure 4. Therefore the number of permutations that lead to the diagonal path T is $1 \cdot 1 \cdot 5 \cdot 2^2 \cdot 1^2 = 20$, and hence the probability of T is $20/5! = 1/6$. $\square$

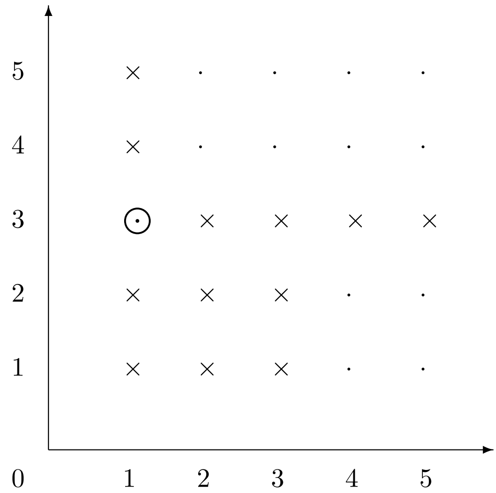

Fig. 2.

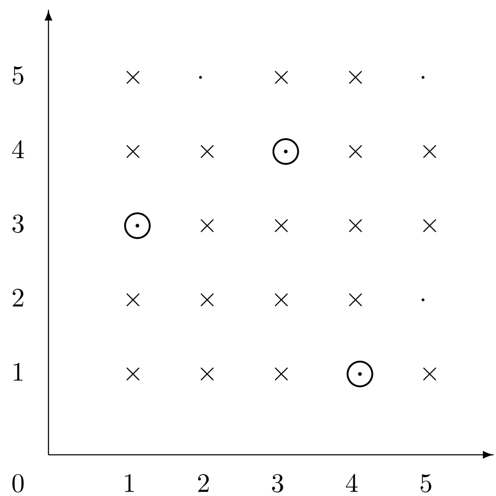

Fig. 3.

3.THE THREE–DIMENSIONAL CASE AND FURTHER

If C is an m-dimensional copula as defined by Schweizer and Sklar [25], then for every $(u_1, \ldots, u_m)$ in $[0,1]^m$ we have that the Fréchet–Hoeffding bounds are as follows:

$$\max(u_1 + \cdots + u_m - m + 1\,,\, 0) \;\leq\; C(u_1, \ldots, u_m) \;\leq\; \min(u_1, \ldots, u_m)\,, \quad (9)$$

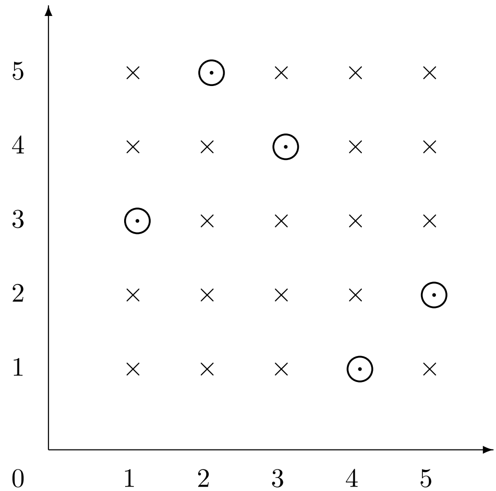

Fig. 4.

but the Fréchet–Hoeffding lower bound is never a copula for $m > 2$, and the above inequality cannot be improved, see Nelsen [24]. According to (9) we have that the diagonal section of an m-copula satisfies

$$\max(mu - m + 1, 0) \leq \delta(u) \leq u, \qquad u \in [0,1].$$

Particularly, the product (or independence) m-copula $\Pi^{(m)}(u_1, \ldots, u_m) = u_1 u_2 \cdots u_m$ has a diagonal section $\delta_\Pi(u) = u^m$. For an Archimedean m-copula and every $m \geq 2$, from Kimberling [14] we have that its generator must be strict and completely monotonic, but this condition is not necessary if $m \geq 3$ and m is *fixed*, and a weaker condition, that is m-monotonicity gives the same result, see McNeil and Nešlehová [22]. But in both cases we have the following expression for its diagonal section: $\delta(u) = \varphi^{-1}\big(m\varphi(u)\big)$, $u \in [0,1]^m$, or equivalently

$$\varphi\big(\delta(u)\big) = m\varphi(u),$$

which again leads us to Schröder's functional equation, see (2). As a particular case of Theorem 6.6 in Kuczma [19] (or Theorem 2.3.12 in Kuczma et al. [20]), let the function $\gamma : [0,1] \to [0,1]$ be such that $0 < \gamma(u) < u$ for all $u \in \,]0,1[$, and $\gamma'(0) = \frac{1}{m}$. If $s(u)$ is a solution of the functional equation $s\big(\gamma(u)\big) = \frac{1}{m}s(u)$ such that the function $s(u)/u$ is monotonic in $]0,1[$, then $s(u) = k\lim_{r\to\infty} m^r\gamma^r(u)$, where γ^r is the r*th iteration* of γ, that is the composition of γ with itself r times, and k any constant. And by a similar argument as in Frank's Theorem, we have that if C is an Archimedean m-copula whose diagonal δ satisfies $\delta'(1-) = m$ then it is uniquely determined by its diagonal. This is the case, for example, of the product m-copula, which in the context of a m-dimensional random vector of continuous random variables represents independence.

We will now analyze analogous properties of the empirical diagonal as done in the previous section, for the case $m = 3$, hoping that this suffices to convince the reader that analogous results may be obtained for higher dimensions.

Let $S := \{(X_1, Y_1, Z_1), \ldots, (X_n, Y_n, Z_n)\}$ denote a random sample of size n from a random vector of continuous random variables (X, Y, Z). As defined by Deheuvels [9], the *trivariate empirical copula* is the function C_n given by

$$C_n\left(\frac{i}{n}, \frac{j}{n}, \frac{k}{n}\right) = \frac{1}{n} \sum_{(X,Y,Z)\in S} \mathbf{1}_{]-\infty\,,\,X_{(i)}]\times]-\infty\,,\,Y_{(j)}]\times]-\infty\,,\,Z_{(k)}]}(X, Y, Z)\,,$$

where $X_{(i)}$, $Y_{(j)}$ and $Z_{(k)}$ denote the order statistics of the sample, for i, j and k in $\{1, \ldots, n\}$, and $C_n(x, y, z) = 0$, whenever any of x, y or z equals 0.

Definition 3.1. The trivariate empirical diagonal is the function δ_n given by

$$\delta_n\left(\frac{j}{n}\right) := C_n\left(\frac{j}{n}, \frac{j}{n}, \frac{j}{n}\right) \quad j = 0, 1, \ldots, n\,.$$

Without loss of generality we may assume that the X_k values in S are ordered, then

$$\delta_n\left(\frac{j}{n}\right) = \frac{1}{n} \sum_{k=1}^{j} \mathbf{1}_{]-\infty\,,\,Y_{(j)}]\times]-\infty\,,\,Z_{(j)}]}(Y_k, Z_k)\,,$$

where $j = 1, \ldots, n-1$, and $\delta_n(0) = 0$, $\delta_n(1) = 1$.

It is clear from above that δ_n is a nondecreasing function of j. Moreover, by Fréchet–Hoeffding bounds:

$$\max\left(\frac{3j}{n} - 2, 0\right) \le \delta_n\left(\frac{j}{n}\right) \le \frac{j}{n}\,. \tag{10}$$

Proposition 3.2.

$$\delta_n\left(\frac{j+1}{n}\right) - \delta_n\left(\frac{j}{n}\right) \in \left\{0, \frac{1}{n}, \frac{2}{n}, \frac{3}{n}\right\}.$$

Proof. Define the random set $A(j) :=]-\infty, Y_{(j)}] \times]-\infty, Z_{(j)}]$. Then

$$\begin{aligned} n\left[\delta_n\left(\frac{j+1}{n}\right) - \delta_n\left(\frac{j}{n}\right)\right] &= \sum_{k=1}^{j+1} \mathbf{1}_{A(j+1)}(Y_k, Z_k) - \sum_{k=1}^{j} \mathbf{1}_{A(j)}(Y_k, Z_k)\,, \\ &= \mathbf{1}_{A(j+1)}(Y_{j+1}, Z_{j+1}) \\ &\quad + \sum_{k=1}^{j} \Big[\mathbf{1}_{A(j+1)}(Y_k, Z_k) - \mathbf{1}_{A(j)}(Y_k, Z_k)\Big]\,, \\ &= \mathbf{1}_{A(j+1)}(Y_{j+1}, Z_{j+1}) + \sum_{k=1}^{j} \mathbf{1}_{A(j+1)\setminus A(j)}(Y_k, Z_k)\,. \end{aligned}$$

Since the last two indicator functions may take values 0 or 1 independently, and since the set $A(j+1)\setminus A(j)$ may contain 0, 1 or 2 points, the result follows. □

This means that all the possible paths $\{\delta_n(\frac{j}{n}) : j=0,1,\ldots,n\}$ are between the paths $\{\max(\frac{3j}{n}-2,0) : j=0,1,\ldots,n\}$ and $\{\frac{j}{n} : j=0,1,\ldots,n\}$ with jumps of size $0, \frac{1}{n}, \frac{2}{n}$ or $\frac{3}{n}$ between consecutive steps.

Definition 3.3. Let $(X_1,Y_1,Z_1),\ldots,(X_n,Y_n,Z_n)$ be a random sample of the random vector (X,Y,Z) of continuous random variables, where $X_1<X_2<\cdots<X_n$. Let us define the trivariate diagonal random path by the vector

$$\boldsymbol{T} = \left(\delta_n\left(\frac{0}{n}\right), \delta_n\left(\frac{1}{n}\right), \ldots, \delta_n\left(\frac{n}{n}\right)\right)$$

where

$$\delta_n\left(\frac{j}{n}\right) := \frac{1}{n}\sum_{k=1}^{j}\mathbf{1}_{A(j)}(Y_k,Z_k)\,, \quad j=1,\ldots,n-1\,,$$

with $\delta_n(0)=0$, $\delta_n(1)=1$.

Associated to the diagonal random path we may define the *diagonal random increments* by the vector $\boldsymbol{I} := \left(\delta_n\left(\frac{1}{n}\right)-\delta_n\left(\frac{0}{n}\right), \delta_n\left(\frac{2}{n}\right)-\delta_n\left(\frac{1}{n}\right), \ldots, \delta_n\left(\frac{n}{n}\right)-\delta_n\left(\frac{n-1}{n}\right)\right)$ so that knowledge of $\boldsymbol{T}=(t_0=0,t_1,\ldots,t_{n-1},t_n=1)$ is equivalent to knowledge of $\boldsymbol{I}=(i_1,\ldots,i_n)$, where $i_j=t_j-t_{j-1}$ and $t_j=\sum_{k=1}^{j} i_k$.

Alternatively, we may write $\boldsymbol{I}=(i_1,\ldots,i_n)=\frac{1}{n}(b_1,\ldots,b_n)$ where the $b_j\in\{0,1,2,3\}$. Moreover, $i_n=1-t_{n-1}=1-\sum_{k=1}^{n-1} i_k$ so knowledge of $\boldsymbol{B}:=(b_1,\ldots,b_{n-1})$ completely specifies any path. Different values of $\boldsymbol{B}$ can be labeled as vectors of base-4 numbers. For example, with $n=7$ the Fréchet–Hoeffding lower bound path $\{\max(\frac{3j}{n}-2,0) : j=0,1,\ldots,n\}$ is specified by the vector $(0,0,0,0,1,3)$, while the Fréchet–Hoeffding upper bound path $\{\frac{j}{n} : j=0,1,\ldots,n\}$ is specified by the vector $(1,1,1,1,1,1)$.

But not every base-4 representation will generate a valid path. For example, for $n=7$ we have that $(0,3,0,0,2,0)$ is a base-4 number between $(0,0,0,0,1,3)$ and $(1,1,1,1,1,1)$, but it represents a path that is out of Fréchet–Hoeffding bounds. In general, we just have to check which of the base-4 representations satisfy

$$\max\left(\frac{3j}{n}-2,0\right) \le t_j \le \frac{j}{n}\,, \quad j=1,\ldots,n-1\,, \tag{11}$$

which is equivalent to satisfy

$$\max(3j-2n,0) \le \sum_{k=1}^{j} b_k \le j\,, \quad j=1,\ldots,n-1\,. \tag{12}$$

We will call an *admissible diagonal path* any vector of base-4 numbers satisfying the Fréchet–Hoeffding conditions (10), or equivalently, conditions (12). In the the following result we find the exact number of admissible paths for any $n\ge 3$.

Proposition 3.4. Let P_n denote the number of admissible paths for the trivariate empirical diagonal δ_n of n points in $[0,1]^3$ and $n \geq 3$. Then

$$P_n = \sum_{r_0=0}^{[\![2n/3]\!]} \sum_{r_2+2r_3=r_0} \frac{\binom{n}{r_0}\binom{n-r_0}{r_3}\binom{n-r_0-r_3}{r_2}}{r_0+1} = \sum_{r_0=0}^{[\![2n/3]\!]} \sum_{r_3=0}^{[\![r_0/2]\!]} \frac{\binom{n}{r_0}\binom{n-r_0}{r_3}\binom{n-r_0-r_3}{r_0-2r_3}}{r_0+1},$$

where $[\![x]\!]$ denotes the greatest integer less than or equal to x.

Proof. Let $(x_1, y_1, z_1), \ldots, (x_n, y_n, z_n) \in [0,1]^3$ be n points and let δ_n be their empirical diagonal. Let r_0, r_1, r_2 and r_3 denote the number of zeros, ones, twos, and threes, respectively, in the empirical diagonal $n\delta_n$. Then $r_0 + r_1 + r_2 + r_3 = n$ and $r_1 + 2r_2 + 3r_3 = n$, so $r_2 + 2r_3 = r_0$. Hence, any nonnegative integers r_1, r_0, r_3 satisfying

$$\begin{aligned} r_1 + 2r_0 - r_3 &= n\,, \\ r_0 &\geq 2r_3\,, \end{aligned} \tag{13}$$

could provide an admissible path for the empirical diagonal whenever the Fréchet–Hoeffding bounds are satisfied. Observe that $r_0 \leq [\![2n/3]\!]$, by (12), and that (13) is a consequence of the fact that $0 \leq r_2 = r_0 - 2r_3$.

If $\delta_n = \frac{1}{n}(b_1, b_2, \ldots, b_n)$ is an admissible path, then $b_1 \in \{0,1\}$, $b_2 \in \{0,1,2\}$, and $b_i \in \{0,1,2,3\}$ for $3 \leq i \leq n$. We observe that the restrictions $\sum_{i=1}^{n} b_i = n$ and $\sum_{i=1}^{j} b_i \leq j$ must be fulfilled. Since $r_0 = r_2 + 2r_3$ now the basic rule to find admissible paths is "one zero before each two, two zeros before each three". Namely, if some $b_i = 2$, then there exists $1 \leq j < i$ such that $b_j = 0$, and if some $b_i = 3$, then there exist $1 \leq j < k < i$ such that $b_j = 0 = b_k$. Therefore, given the number of zeros and threes among the b_i's, we only have to see where can they be located in the vector $(b_1, \ldots, b_n)$, following the basic rule, with $r_2 = r_0 - 2r_3$. Observe that the ones can be located in any place.

First assume that $r_0 = 0$, then (13) implies $r_3 = 0$, and $r_2 = r_0 - 2r_3 = 0$, so the only possibility is $r_1 = n$, that is, the admissible path $(1, 1, \ldots, 1)$, which is Fréchet–Hoeffding upper bound path. Then, by an analogous argument as in Proposition 2.3, for any r_0 positive integer such that $r_0 \leq [\![2n/3]\!]$, and any nonnegative integer r_3 such that $r_2 + 2r_3 = r_0$, the number of admissible paths with r_0 zeros and r_3 threes is given by

$$\frac{\binom{n}{r_0}\binom{n-r_0}{r_3}\binom{n-r_0-r_3}{r_2}}{r_0+1} = \frac{1}{r_0+1}\binom{n}{r_0,\ r_1,\ r_2,\ r_3}, \tag{14}$$

where the right side of this last equation is justified by analogous arguments as in the Remark of Proposition 2.3. The result now follows summing over all possible values of r_0 and r_3, subject to the constraint $r_2 + 2r_3 = r_0$, which is equivalent to sum

$$\frac{\binom{n}{r_0}\binom{n-r_0}{r_3}\binom{n-r_0-r_3}{r_0-2r_3}}{r_0+1},$$

over all possible values of $r_0 \leq [\![2n/3]\!]$, and $r_3 \leq [\![r_0/2]\!]$, by (13). $\square$

Remark. A thorough justification for (14) may be given in terms of Riordan Group enumeration techniques, as in Cameron [6].

Now we will calculate the probability of any given (admissible) path, under the hypothesis of independence.

Theorem 3.5. Let $S = \{(X_1, Y_1, Z_1), \ldots, (X_n, Y_n, Z_n)\}$ be a random sample from the random vector of continuous random variables (X, Y, Z). If X, Y and Z are independent and if $\boldsymbol{T} = (t_0 = 0, t_1, \ldots, t_{n-1}, t_n = 1)$ is an admissible diagonal path, then

$$\Pr\left[\, \boldsymbol{T} = (t_0 = 0, t_1, \ldots, t_{n-1}, t_n = 1) \,\right] = \frac{1}{(n!)^2} \prod_{j=1}^{n} f(j)\,,$$

where $f(j)$ is obtained in terms of the following formula, for $j = 1, \ldots, n$:

$$f(j) = \begin{cases} 1 & \text{if} \quad n(t_j - t_{j-1}) = 0\,, \\ 3(j - nt_{j-1})(j - 1 - nt_{j-1}) + 1 & \text{if} \quad n(t_j - t_{j-1}) = 1\,, \\ 3(j - 1 - nt_{j-1})^4 & \text{if} \quad n(t_j - t_{j-1}) = 2\,, \\ (j - 1 - nt_{j-1})^3 (j - 2 - nt_{j-1})^3 & \text{if} \quad n(t_j - t_{j-1}) = 3\,. \end{cases}$$

Proof. From the continuity assumption we know that, with probability one, there are no ties among the X_i's, the Y_i's or the Z_i's. By independence of X, Y and Z, the probability of the random sample $\{(X_1, Y_1, Z_1), \ldots, (X_n, Y_n, Z_n))\}$ equals that of $\{(X_{(1)}, Y_{\sigma(1)}, Z_{\tau(1)}), \ldots, (X_{(n)}, Y_{\sigma(n)}, Z_{\tau(n)})\}$ where $(\sigma(1), \tau(1)), \ldots, (\sigma(n), \tau(n))$ is any bivariate permutation of $I_n := \{1, \ldots, n\}$, and every permutation has probability $(n!)^{-2}$.

By rescaling we can assume that $X_{(i)} = i$, $Y_{\sigma(i)} = \sigma(i)$, and $Z_{\tau(i)} = \tau(i)$, for $i \in I_n$, that is to consider the one-to-one rank-mapping $(X_i, Y_j, Z_k) \mapsto (rank(X_i), \sigma(i), \tau(i))$. Hence the rank-mapped sample S becomes a subset of the (three-dimensional) grid $\{1, 2, \ldots, n\}^3 := I_n^3$. In fact, for every $i \in I_n$ there exists a unique $(j, k) = (\sigma(i), \tau(i)) \in I_n^2$ such that $(i, j, k) = (i, \sigma(i), \tau(i)) \in S$. That is, for any horizontal or vertical segment in the three bivariate grids $\{i\} \times I_n^2$, $I_n \times \{j\} \times I_n$, and $I_n^2 \times \{k\}$, there is exactly one point that belongs to the sample S.

In order to calculate $\Pr\left[\, \boldsymbol{T} = (0, t_1, \ldots, t_{n-1}, 1) \,\right]$ we just need to count the number of orderings of $\{Y_1, \ldots, Y_n\}$ and $\{Z_1, \ldots, Z_n\}$ that would lead to the admissible path $(t_0 = 0, t_1, \ldots, t_{n-1}, t_n = 1)$. We will show that this probability is given by $(n!)^{-2} \prod_{j=1}^{n} f(j)$.

Let $\boldsymbol{B} = (b_1, b_2, \ldots, b_n)$ be the vector of base-4 numbers which is equivalent to the admissible diagonal path $\boldsymbol{T} = (t_0 = 0, t_1, \ldots, t_{n-1}, t_n = 1)$, that is $b_i = n(t_i - t_{i-1})$ for $i = 1, 2, \ldots, n$.

If $b_k = 0 = n(t_k - t_{k-1})$ then there is only one possibility: no sample point of S is rank-mapped to $\{1, \ldots, k\}^3 \setminus \{1, \ldots, k-1\}^3$.

Define

$$K := \min\{i \in \{1, \ldots, n\} \,|\, b_i > 0\}.$$

Since $\mathbf{B}$ represents an admissible diagonal path, we have that all the $b_i\,'s$ are equal to $0,1,2$ or 3, except for b_1 which equals 0 or 1, and b_2 which equals $0,1$ or 2, and $\sum_{i=1}^{n} b_i = n$. Then $K \leq [\![\, 2n/3 \,]\!]$ by (11).

If $K = 1$ it means that $b_1 = 1$, and then $(1,1,1) \in S$, and there is only one possibility for $\sigma(1)$ and $\tau(1)$, that is $(\sigma(1), \tau(1)) = (1,1)$.

Define $\mathcal{D}_1 := \{(1,1,1)\}$ and for $r = 2,3,\ldots$ let $\mathcal{D}_r := \{1,2,\ldots,r\}^3 \setminus \bigcup_{w=1}^{r-1} \mathcal{D}_w$. Then $\mathrm{card}(\mathcal{D}_r) = r^3 - (r-1)^3 = 3\,r(r-1)+1$, where $\mathrm{card}(\,\cdot\,)$ stands for cardinality of a set. Geometrically, $\mathcal{D}_r$ may be interpreted as a grid on three faces of a cube of volume r^3, with one *vertex,* $3(r-1)$ points on the three edges (excluding the vertex), which we will call *edge points*, and therefore

$$\frac{r^3 - (r-1)^3 - 3(r-1) - 1}{3} = (r-1)^2$$

points on each face (without edges), which we will call *face points.* So in $\mathcal{D}_r$ we have $3(r-1)^2$ face points, $3(r-1)$ edge points, and exactly 1 vertex. All points $(i,j,k) \in \mathcal{D}_r$ must have at least one entry equal to r. If a point in $\mathcal{D}_r$ has only one entry equal to r then it is a face point; if it has 2 entries equal to r and the other one different from r then it is an edge point; and if it has its 3 entries equal to r it is obviously the vertex of $\mathcal{D}_r$.

If $K = 2$ it means that $b_1 = 0 = nt_1$, that is $(1,1,1) \notin S$, and $b_2 \in \{1,2\}$. In case $b_2 = 1$, there is only one point $(X_i, Y_j, Z_k) \in S$ which is rank-mapped to one of the elements of $\mathcal{D}_2$, that is, there are $\mathrm{card}(\mathcal{D}_2) = 2^3 - 1^3 = 7$ possibilities for such point. In case $b_2 = 2$, there are exactly two points $(X_i, Y_j, Z_k) \in S$ which are rank-mapped to 2 different elements of $\mathcal{D}_2$, and the number of possibilities depends on whether one of the points belongs to one of the $3(2-1)+1 = 4$ points that lie on the 3 edges of $\mathcal{D}_2$. First of all, we have to discard the vertex $(2,2,2)$ since this point belongs to the three edges of $\mathcal{D}_2$, and this would eliminate the possibility of using any other point in the three faces of $\mathcal{D}_2$, and we need to allocate two points. If one of the points is an edge point, then it automatically eliminates the possibility of choosing the other point from 2 faces and the 3 edges of $\mathcal{D}_2$, that is, the other point has to be a face point, so at least one of the two points has to be a face point. So first we count the number of ways in which we can choose the face point, which is $3 : (1,1,2)$, $(1,2,1)$, $(2,1,1)$; then, its selection eliminates 2^2 points on the face where it is located (the vertex included) plus $(2-1) = 1$ face points on each of the other two faces, and so there is left $7 - 2^2 - 2(1) = 1$ possibility for the other point, for a total of $3(1) = 3$ different ways of choosing the two points.

Now assume that $K \geq 3$, and therefore $b_K \in \{1,2,3\}$. This implies that $b_1 = \cdots = b_{K-1} = 0$, which is equivalent to $t_1 = \cdots = t_{K-1} = 0$, that is, there are no points in the sample S which are rank-mapped to the set $\bigcup_{w=1}^{K-1} \mathcal{D}_w = \{1,\ldots,K-1\}^3$. From the definition of the trivariate empirical copula it is clear that

$$nt_i = nC_n\left(\frac{i}{n}, \frac{i}{n}, \frac{i}{n}\right) = \mathrm{card}(S \cap \{1,\ldots,i\}^3)\,,$$

that is, nt_i is the number of (rank-mapped) sample points in S that belong to $\{1,\ldots,i\}^3$.

If $b_K = 1$, or equivalently $nt_K = 1$, and since $nt_{K-1} = 0$ implies that there are not any points in S rank-mapped to $\{1, \ldots, K-1\}^3$, we have that there is exactly one point of S which is rank-mapped to $\mathcal{D}_K$, and there are card$(\mathcal{D}_K) = K^3 - (K-1)^3 = 3K(K-1)+1$ different possibilities to choose this point. It is important to mention that the corresponding rank-mapped point, say (i^*, j^*, k^*), automatically cancels the possibility that any other point of the sample S is rank-mapped to a point $(i, j, k) \in \{1, \ldots, n\}^3$ such that $i = i^*$ or $j = j^*$ or $k = k^*$.

If $b_K = 2$ then there are exactly 2 sample points of S which are rank-mapped to 2 different elements of $\mathcal{D}_K$, and the number of possibilities depends on whether one of the points belongs to one of the $3(K-1)+1$ points that lie on the 3 edges of $\mathcal{D}_K$. First of all, we have to discard the vertex (K, K, K) since this point belongs to the three edges of $\mathcal{D}_K$, and this would eliminate the possibility of using any other point in the three faces of $\mathcal{D}_K$, and we need to allocate two points. If one of the points is an edge point then it automatically eliminates the possibility of choosing the other point from 2 faces and the 3 edges of $\mathcal{D}_K$, that is, the other point has to be a face point, so at least one of the two points has to be a face point. So first we count the number of ways in which we can choose the face point, which is $\binom{3}{1}(K-1)^2$; then, its selection eliminates K^2 points on the face where it is located, and so it eliminates $2K-1$ points of the K^2 points on each of the other two faces, that is, $K^2 - (2K-1) = (K-1)^2$ points are left on each of the other two faces; we may choose one out of the two faces left and so we have $\binom{2}{1}(K-1)^2$ different possibilities for the second point, and so we have a total of

$$\frac{\binom{3}{1}(K-1)^2\binom{2}{1}(K-1)^2}{2!} = 3(K-1)^4$$

different ways of choosing the two points (we divided by 2! since the order of the two points chosen is not important).

If $b_K = 3$ then there are exactly 3 sample points of S which are rank-mapped to 3 different elements of $\mathcal{D}_K$, which necessarily have to be face points (one on each of the three faces of $\mathcal{D}_K$) since the presence of one edge point would just leave one (or zero in the case of the vertex) faces for choosing the other two points, which is impossible since it is only possible to have one point per face. Then, we have $\binom{3}{1}(K-1)^2$ different ways of choosing the first point, which just leaves available $(K-1)^2 - (K-1) = (K-1)(K-2)$ points on each of the other two faces, so we may choose the second point in $\binom{2}{1}(K-1)(K-2)$ different ways, which in turn will eliminate $(K-2)$ points of the remaining face, leaving $\binom{1}{1}[(K-1)(K-2) - (K-2)] = \binom{1}{1}(K-2)^2$ ways of choosing the third point, for a total of

$$\frac{\binom{3}{1}(K-1)^2\binom{2}{1}(K-1)(K-2)\binom{1}{1}(K-2)^2}{3!} = [(K-1)(K-2)]^3$$

different ways for choosing the 3 points (we divided by 3! since the order of the three points chosen is not important).

For b_J's with $J > K$ we have that $b_J \in \{0, 1, 2, 3\}$ and we proceed in an analogous way, but eliminating the points $(i, j, k) \in \mathcal{D}_J$ for which there exists $(i^*, j^*, k^*) \in$

$\bigcup_{w=1}^{J-1} \mathcal{D}_w = \{1, \ldots, J-1\}^3$ such that $i = i^*$ or $j = j^*$ or $k = k^*$. For the calculations we proceed in an analogous ways as for $j = K$ by just eliminating for each of the three dimensions nt_{J-1} points since nt_{J-1} is the number of (rank-mapped) sample points in S that belong to $\{1, \ldots, J-1\}^3$, and so we arrive to the same formulas by just replacing K with $J - nt_{J-1}$ and the result follows. □

4.FINAL REMARKS

In the present work, for $m = 2$ and $m = 3$, we have proved that it is possible to:

- label the different paths that a m-variate empirical diagonal may follow by using base-$(m+1)$ number representation,
- count the number of admissible diagonal paths, given a sample size n (Propositions 2.3 and 3.4),
- obtain the exact distribution of the empirical diagonal under the hypothesis of independence of a vector of continuous random variables (Theorems 2.4 and 3.5),

with the possibility of obtaining analogous results for higher dimensions.

The results in the present work are useful to give solution to an open problem proposed in Alsina, Frank and Schweizer [2]:

> Can one design a test of statistical independence based on the assumptions that the copula in question is Archimedean and that its diagonal section is $\delta(u) = u^2$?

If we are interested in analyzing independence of two continuous random variables, the results stated in the Introduction suggest to measure some kind of closeness between the empirical diagonal and the diagonal section of the product copula. Moreover, a nonparametric test of independence can be carried out, as suggested by Sungur and Yang [27], using the diagonal. Let (X, Y) be a random vector of continuous random variables with Archimedean copula C, then the following hypothesis are equivalent:

$$\mathcal{H}_0 : X \text{ and } Y \text{ are independent} \quad \Leftrightarrow \quad \mathcal{H}_0^* : C = \Pi \quad \Leftrightarrow \quad \mathcal{H}_0^{**} : \delta_C(u) = u^2 . \tag{15}$$

Using the results of the previous sections, we may propose a statistical test based on the empirical diagonal because under $\mathcal{H}_0$ we know the exact distribution of the empirical diagonal (Theorem 2.4) and so we can obtain the *exact distribution* of any test statistic based on it, and therefore such a test will be also useful under small samples, in contrast with other nonparametric tests which rely on the asymptotic behavior of their proposed statistics. For example, a first idea would be to work with a Cramér–von Mises-type test statistic based on the empirical diagonal:

$$CvM_n := \frac{1}{n-1} \sum_{j=1}^{n-1} \left(\delta_n\left(\frac{j}{n}\right) - \frac{j^2}{n^2} \right)^2 , \tag{16}$$

rejecting $\mathcal{H}_0$ whenever CvM_n is "too large," that is, if $CvM_n \geq k_\alpha$ for $0 < \alpha < 1$ a given test size, where $k_\alpha = \min\{\, x : \Pr[\, CvM_n \geq x \,|\, \mathcal{H}_0\,] \leq \alpha\,\}$. An analogous test may be carried out for a random vector of continuous random variables (X, Y, Z) with Archimedean copula C, using Theorem 3.5 in this work.

Any other statistic based on the empirical diagonal behavior, could be defined and its exact distribution could be found using Theorems 2.4 and 3.5. The study of the power of any test based on such statistics are beyond the scope of this paper.

Some of these results appeared in Erdely [11].

ACKNOWLEDGEMENT

We are specially thankful to Roger B. Nelsen and Radko Mesiar for useful comments about references and nice counterexamples. We are also thankful to two anonymous referees whose indications helped to underline the primary goal and main results of the present work and for relevant references. This work was partially supported by Mexico's CONACyT grant 50152.

(Received November 13, 2007.)

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] D. F. Bailey: Counting arrangements of 1's and −1's. Math. Mag. *69* (1996), 128–131.

[4] E. Barcucci and M. C. Verri: Some more properties of Catalan numbers. Discrete Math. *102* (1992), 229–237.

[5] D. Callan: Some bijections and identities for the Catalan and Fine numbers. Séminaire Lotharingen de Combinatoire *53* (2006), B53e (16 pp).

[6] N. T. Cameron: Random Walks, Trees and Extensions of Riordan Group Techniques. Ph.D. Thesis. Howard University 2002.

[7] K. L. Chung and W. Feller: On fluctuations in coin-tossing. Proc. Nat. Acad. Sci. U. S. A. *35* (1949), 605–608.

[8] W. F. Darsow and M. J. Frank: Associative functions and Abel–Schröder systems. Publ. Math. Debrecen *30* (1983), 253–272.

[9] 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.

[10] E. O. Engleberg: On some problems concerning a restricted random walk. J. Appl. Probab. *2* (1965), 396–404.

[11] A. Erdely: Diagonal Properties of the Empirical Copula and Applications. Construction of Families of Copulas with Given Restrictions. Ph.D. Thesis, Universidad Nacional Autónoma de México 2007.

[12] W. Feller: An Introduction to Probability Theory and Its Applications, Vol. 1. Third edition. Wiley, New York 1968.

[13] M. J. Frank: Diagonals of copulas and Schröder's equation. Aequationes Math. *51* (1996), 150.

[14] C. H. Kimberling: A probabilistic interpretation of complete monotonicity. Aequationes Math. *10* (1974), 152–164.

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

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

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

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

[19] M. Kuczma: Functional Equations in a Single Variable. Polish Scientific Publishers, Warsaw 1968.

[20] M. Kuczma, B. Choczewski, and R. Ger: Iterative Functional Equations. Cambridge University Press, New York 1990.

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

[22] A. J. McNeil and J. Nešlehová: Multivariate Archimedean copulas, D-monotone functions and ℓ_1-norm symmetric distributions. Ann. Statist, to appear.

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

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

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

[26] A. Sklar: Fonctions de répartition à n dimensions et leurs marges. Publ. Inst. Statist. Univ. Paris *8* (1959), 229–231.

[27] E. A. Sungur and Y. Yang: Diagonal copulas of Archimedean class. Comm. Statist. Theory Methods *25* (1996), 1659–1676.

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.

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