A Small-Sample Nonparametric Independence Test for the Archimedean Family of Bivariate Copulas

Arturo Erdely[1] and José M. González-Barrios[2]

[1] Escuela de Actuaría, Universidad Anáhuac México Norte, Huixquilucan, México
[2] Departamento de Probabilidad y Estadística, Instituto de Investigaciones en Matemáticas Aplicadas y en Sistemas, Universidad Nacional Autónoma de México, México D.F., México

Abstract. In this paper we study the problem of independence of two continuous random variables using the fact that there exists a unique copula that characterizes independence, and that such copula is of Archimedean type. We use properties of the empirical diagonal to build nonparametric independence tests for small samples, under the assumption that the underlying copula belongs to the Archimedean family, giving solution to an open problem proposed by Alsina et al. [2].

1 Introduction

A *bivariate copula* is a function $C : [0,1]^2 \to [0,1]$ with the following properties: For every u,v in $[0,1]$, $C(u,0) = 0 = C(0,v)$, $C(u,1) = u$ and $C(1,v) = v$, and for every u_1,u_2,v_1,v_2 in $[0,1]$ such that $u_1 \leq u_2$ and $v_1 \leq v_2$, $C(u_2,v_2) - C(u_2,v_1) - C(u_1,v_2) + C(u_1,v_1) \geq 0$. Also, $W(u,v) \leq C(u,v) \leq M(u,v)$, where $W(u,v) := \max(u+v-1,0)$ and $M(u,v) := \min(u,v)$, where W and M are themselves copulas, known as the *Fréchet-Hoeffding lower and upper bounds,* respectively. The *diagonal section* of a bivariate copula, $\delta_C(u) := C(u,u)$, is a nondecreasing and uniformly continuous function on $[0,1]$ where: i) $\delta_C(0) = 0$ and $\delta_C(1) = 1$; ii) $0 \leq \delta_C(u_2) - \delta_C(u_1) \leq 2(u_2 - u_1)$ for all u_1,u_2 in $[0,1]$ with $u_1 \leq u_2$; iii) $\max(2u-1,0) \leq \delta_C(u) \leq u$. A copula C is said to be *Archimedean*, see [17], if $C(u,v) = \varphi^{[-1]}[\varphi(u) + \varphi(v)]$, where φ, called the *generator* of the copula, 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 given by: $\varphi^{[-1]}(t) := \varphi^{-1}(t)$ if $0 \leq t \leq \varphi(0)$, and $\varphi^{[-1]}(t) := 0$ if $\varphi(0) \leq t \leq \infty$. Its diagonal section is given by $\delta_C(u) = \varphi^{[-1]}[2\varphi(u)]$. One may ask, as observed in [6], given δ, what can be said about φ? The following result is part of what was proved in [9] and [3]:

Theorem 1. *If C is an Archimedean copula whose diagonal δ satisfies $\delta'(1-) = 2$ then C is uniquely determined by its diagonal.*

From now on we will refer to the condition $\delta'(1-) = 2$ as *Frank's condition.* 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 [20], and so it is uniquely determined by its diagonal section $\delta_\Pi(u) = u^2$. Frank's condition is satisfied by 13 out of 22 copulas in the catalog of Archimedean copulas provided by [17].

D. Dubois et al. (Eds.): Soft Methods for Hand. Var. and Imprecision, ASC 48, pp. 118–125, 2008.
springerlink.com

2 The Empirical Diagonal and Some Properties

In the case of Archimedean bivariate copulas, the diagonal section contains all the information we need to build the copula, provided that Frank's condition $\delta'(1-) = 2$ is satisfied, 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.

Let $S := \{(x_1, y_1), \ldots, (x_n, y_n)\}$ denote a sample of size n from a continuous random vector (X, Y). The *empirical copula* is the function C_n given by (see [17])

$$C_n\left(\frac{i}{n}, \frac{j}{n}\right) = \frac{1}{n} \sum_{k=1}^{n} \mathbf{1}\{x_k \leq x_{(i)}, y_k \leq y_{(j)}\},$$

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 $\{0, 1/n, \ldots, (n-1)/n, 1\}$.

Remark 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 [15] and [16]. Moreover, an empirical copula is an example of an irreducible discrete copula as defined in [13]. An empirical copula is not a copula, but a (two-dimensional) *subcopula*, for details of subcopulas see [17]. We should notice also the following relationship between the empirical copula and the empirical joint distribution function $H_n : C_n(\frac{i}{n}, \frac{j}{n}) = H_n(x_{(i)}, y_{(j)})$.

Definition 1. *The* empirical diagonal *is the function* $\delta_n(j/n) := C_n(j/n, j/n)$ *for* $j = 0, 1, \ldots, n$, *and* $\delta_n(0) := 0$.

It is clear from above that δ_n is a nondecreasing function of j. Moreover, by Fréchet-Hoeffding bounds for subcopulas we have that $\max(2j/n - 1, 0) \leq \delta_n(j/n) \leq j/n$, and it is also straightforward to prove that the difference $\delta_n((j+1)/n) - \delta_n(j/n)$ equals one of the values $\{0, 1/n, 2/n\}$. These properties also follow from properties of the diagonal section in discrete copulas and quasi-copulas, see [1] or [14].

We will call an *admissible diagonal path* any path $\{\delta_n(j/n) : j = 0, 1, \ldots, n\}$ satisfying the Fréchet-Hoeffding bounds, that is any path between the paths $\{\max(2j/n - 1, 0) : j = 0, 1, \ldots, n\}$ and $\{j/n : j = 0, 1, \ldots, n\}$, with jumps of size $0, 1/n$, or $2/n$ between consecutive steps. The proof of the following theorem is in [7]:

Theorem 2. *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* $\mathbf{T} = (t_0 = 0, t_1, \ldots, t_{n-1}, t_n = 1)$ *is an admissible diagonal path, then*

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

where, for $j = 1, \ldots, n$: $f(j) = 1$ *if* $n(t_j - t_{j-1}) = 0$; $f(j) = 2(j - nt_{j-1}) - 1$ *if* $n(t_j - t_{j-1}) = 1$; *and* $f(j) = (j - 1 - nt_{j-1})^2$ *if* $n(t_j - t_{j-1}) = 2$.

3 A Nonparametric Test for Independence under the Archimedean Family of Bivariate Copulas

In this section we give solution to an open problem proposed in [2] and [3]:

> 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 $\delta(u) = u^2$?

As a corollary of Sklar's Theorem, see [20, 19, 17], we know that if X and Y are 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*. Recall that the product copula is Archimedean and it is characterized by the diagonal section $\delta_\Pi(u) = u^2$. If we are interested in analyzing independence of two continuous random variables, the previous results 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 [2, 21], using the diagonal section. Let (X,Y) be a random vector of continuous random variables with Archimedean copula C, then the following hypothesis are equivalent:

$$H_0 : X \text{ and } Y \text{ are independent} \quad \Leftrightarrow \quad H_0^* : C = \Pi \quad \Leftrightarrow \quad H_0^{**} : \delta_C(u) = u^2. \tag{1}$$

Using the results of the previous sections, we wish to propose a statistical test based on the empirical diagonal because under H_0 we know the exact distribution of the empirical diagonal (Theorem 2) and so we could theoretically obtain the **exact distribution** of any test statistic based on it. A first idea would be to work with a Cramér-von Mises type test statistic based on the empirical diagonal:

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

rejecting H_0 whenever $CvM_n \geq k_\alpha$ for α a given test size. The performance of a test based on (2) will be analyzed later in a short simulation study. Under some Archimedean families, a test based on (2) can be improved under certain alternatives by the following idea: It is straightforward to verify that under H_0 the expectation $\mathbf{E}[\delta_n(j/n)] = \delta_\Pi(j/n) = j^2/n^2$ so we define for $j = 1,\ldots,n-1$ the quotient $\xi(j/n) := |\delta_n(j/n) - j^2/n^2| / (j/n - \max(2j/n - 1, 0))$ as a way of measuring pointwise closeness to independence, noticing that the denominator just standardizes dividing by the distance between the Fréchet-Hoeffding bounds at point j/n, in the spirit of a correction as in [4]. It is straightforward to verify that $0 \leq \xi(j/n) \leq \max(j/n, 1 - j/n) \leq 1 - 1/n$. We propose as a test statistic

$$S_n := \frac{1}{n-1}\sum_{j=1}^{n-1}\xi\left(\frac{j}{n}\right), \tag{3}$$

rejecting H_0 whenever $S_n \geq k_1(\alpha)$, for α a given test size. Before we proceed, let us denote by $\delta_M(u) = u$ and $\delta_W(u) = \max(2u-1, 0)$ the upper and lower Fréchet-Hoeffding diagonal bounds, respectively. For u in $[0,1]$ the average distance between

$\delta_\Pi(u)$ and $\delta_M(u)$ is $1/6$ while the average distance between $\delta_\Pi(u)$ and $\delta_W(u)$ is $1/12$, this means that the diagonal that represents independence is, on average, twice closer to the lower than to the upper Fréchet-Hoeffding diagonal bound, thus independence is far from being in the middle of such bounds, and so we should consider the possibility of taking this into account in defining a test statistic. We define $h(j/n) := (j/n - j^2/n^2)\,/\,(j^2/n^2 - \max(2j/n - 1, 0))$ as a factor to be multiplied by $\xi(j/n)$ for those observations for which $\delta_n(j/n) < j^2/n^2$, in order to compensate somehow the non-equal closeness of the independence diagonal to the Fréchet-Hoeffding bounds. In other words, let us define $\nu(j/n) := h(j/n)\xi(j/n)$ if $\delta_n(j/n) < j^2/n^2$, and $\nu(j/n) := \xi(j/n)$ if $\delta_n(j/n) \geq j^2/n^2$.

We have that $h(j/n)$ is symmetric with respect to $1/2$ and that $1 \leq h(j/n) \leq h(1/n) = h(1 - 1/n) = n - 1$. We now propose the following test statistic

$$A_n := \frac{1}{n-1} \sum_{j=1}^{n-1} \nu\left(\frac{j}{n}\right), \tag{4}$$

rejecting H_0 when $A_n \geq k_2(\alpha)$, for α a given test size. The test statistics (3) and (4) alone lead to biased tests of independence, but an appropriate combination of both leads to an approximately unbiased independence test, by using the decision rule

$$\text{reject } H_0 \text{ whenever } S_n \geq k_1 \text{ or } A_n \geq k_2\,, \tag{5}$$

where $\text{Prob}\big(\{S_n \geq k_1\} \cup \{A_n \geq k_2\}\,|\,H_0\big) \leq \alpha$, for k_1 and k_2 chosen appropriately, according to a given test size α. From their definitions it is immediate to verify that $0 < S_n \leq A_n \leq 3/4 - 1/4n$. Even though the election of (k_1, k_2) is not unique, in order to obtain an approximately unbiased test, a good choice for the alternative hypotheses we will consider is (k_1, k_2) such that $\alpha_1 = \Pr(S_n \geq k_1\,|\,H_0) \approx \Pr(A_n \geq k_2\,|\,H_0) = \alpha_2$. We cannot prove this in general for all possible alternatives since the power of the test for $\theta \neq \theta_0$ depends on the distribution under the alternative hypothesis, but it seems to work adequately in the following simulations for the given alternatives.

Since the main goal of the present work is to give solution to the open problem proposed by [2], building the required independence test, we include a short simulation study just to show that the proposed tests work, without pretending that they are extremely powerful, and we made some comparisons against a few well-known independence tests, without pretending that they constitute an exhaustive list of independence tests:

- Spearman's test, see [11].
- The modified Hoeffding test as introduced in [5].
- A test in [12].

The simulated power comparisons presented here were obtained with sample sizes $n = 15, 50$, $\alpha = 0.05$. Every Monte Carlo experiment reported here has been simulated 10,000 times, using some one-parameter Archimedean and Non-Archimedean copulas as alternatives. In both cases we will consider families of copulas $\{C_\theta\}$ with one-dimensional parameter θ such that there exists a unique θ_0 such that $C_{\theta_0} = \Pi$ or $\lim_{\theta \to \theta_0} C_\theta = \Pi$. The null hypothesis (1) becomes $H_0 : \theta = \theta_0$ versus the alternative $H_1 : \theta \neq \theta_0$.

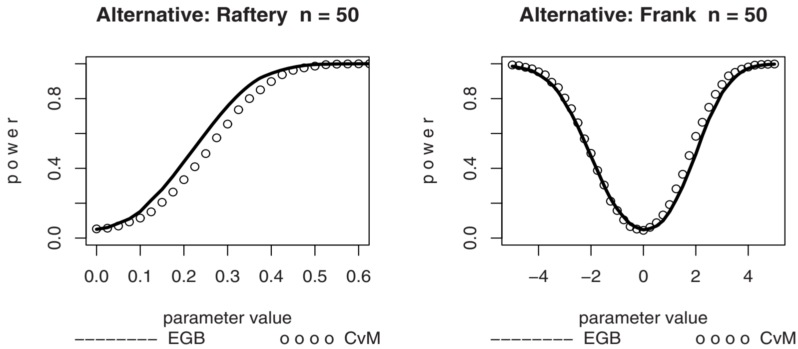

Fig. 1. Left: EGB vs CvM under Raftery. Right: EGB vs CvM under Frank

We will denote by CvM and EGB the tests proposed by the authors in (2) and (5), respectively. Under some families of copulas, there is a clear outperformance of EGB over CvM, for example, with the Raftery family as alternative; but under some other families it is almost the opposite, for example, with the Frank family as alternative, see Fig. 1. The proposed tests EGB and CvM will be compared against the already mentioned tests: R (Spearman), B ([5]), and V ([12]).

Archimedean alternatives. We compared the test powers for $H_0 : \theta = 0$ against $H_1 : \theta \neq 0$ under the following alternative families of Archimedean copulas, for details see [17]: Clayton, Frank, Nelsen's catalog number 4.2.7, Ali-Mikhail-Haq, and Gumbel-Barnett. In all cases these copulas satisfy $C_\theta = \Pi$ if and only if $\theta = 0$, or $\lim_{\theta \to 0} C_\theta = \Pi$, and satisfy Frank's condition $\delta'(1-) = 2$. For example, for the Clayton family see Fig. 2.

Non-Archimedean alternatives. An obvious question is what happens with the proposed EGB and CvM tests outside the Archimedean world. As proved in [10] it is possible to build copulas different from the product (or independence) copula $\Pi(u,v) = uv$

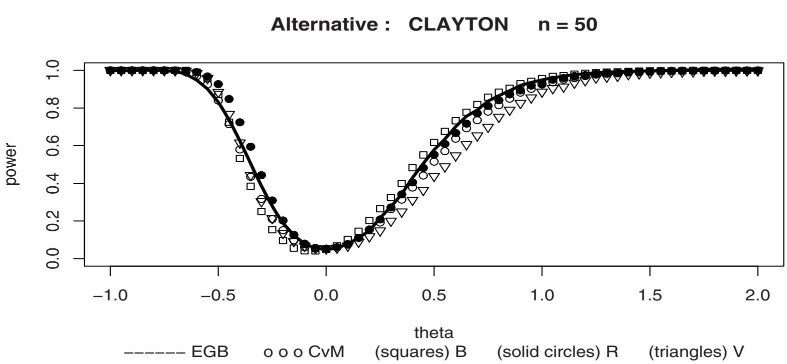

Fig. 2. All tests under Clayton

with the same diagonal as Π, but they are singular, and such copulas rarely appear in real problems. What really might be an issue for the proposed EGB and CvM tests is the fact that there are absolutely continuous non-Archimedean copulas which have the same diagonal as Π, as proved in [8], or as a consequence of the results in [18], so outside the Archimedean world the proposed EGB and CvM tests may face dependence structures that they will not be able to detect. Anyway, we performed similar simulation studies under some well-known non-Archimedean families of copulas, with surprising results. We compared the test powers for $H_0 : \theta = \theta_0$ against $H_1 : \theta \neq \theta_0$ under the following alternative non-Archimedean families of copulas: Raftery, Cuadras-Augé, Farlie-Gumbel-Morgenstern, and Plackett (for details of these families see [17]). In all cases these copulas satisfy $C_\theta = \Pi$ if and only if $\theta = \theta_0$, or $\lim_{\theta \to 0} C_\theta = \Pi$, with $\theta_0 = 0$ for the first three families, and $\theta_0 = 1$ for the last one.

Summary of results. We made a summary of the power comparisons in the format suggested by [12]: For each test statistic, we have calculated the difference between the power of the test and the maximal power of the tests under consideration at the given alternative. For each graph this difference is maximized over the alternatives in the graph. This number can be seen as a summary for the behavior of the test in that graph, although of course some information of the graph is lost. In Table 1 we present percentage differences in maximal power of the five tests under comparison at various alternatives, so that the lower the difference number in the table, the better is the relative performance of the test.

Table 1. Relative power performance

n = 15	Alternative Copula	EGB	CvM	R	B	V
	Clayton	31	43	35	78	50
	Frank	40	37	34	75	54
	Nelsen 4.2.7	36	49	5	77	9
	Ali-Mikhail-Haq	43	37	33	76	55
	Gumbel-Barnett	24	45	13	78	44
	Raftery	19	29	29	5	31
	Cuadras-Augé	25	25	37	0	41
	Farlie-Gumbel-Morgenstern	48	37	32	77	57
	Plackett	42	38	33	73	53

n = 50	Alternative Copula	EGB	CvM	R	B	V
	Clayton	27	32	24	56	44
	Frank	42	27	24	50	52
	Nelsen 4.2.7	28	49	22	70	15
	Ali-Mikhail-Haq	40	28	24	50	53
	Gumbel-Barnett	20	33	8	58	42
	Raftery	4	31	32	20	34
	Cuadras-Augé	12	16	32	8	37
	Farlie-Gumbel-Morgenstern	44	26	25	51	53
	Plackett	40	26	18	43	49

In practice, when using a nonparametric test for independence we usually do not know what alternative we are dealing with, so what is valuable about a test is its ability to maintain an acceptable performance under different alternatives, rather than being the best under specific ones. In this sense, it seems that in general terms, the R test would be the best choice among the tests considered, followed by the EGB and CvM proposed tests.

4 Final Remark

If the underlying copula of a random vector (X,Y) is of the Archimedean type, independence tests can be carried out by defining appropriate test statistics based on the empirical diagonal. Such statistics are discrete random variables and their **exact distribution** may be obtained using Theorem 2, so no asymptotic approximations are required, which may be specially helpful with small samples.

References

1. Aguiló, I., Suñer, J., Torrens, J.: Matrix representation of discrete quasi-copulas. Fuzzy Sets Syst. (in press, 2008) doi:10.1016/j.fss2007.10.004
2. Alsina, C., Frank, M.J., Schweizer, B.: Problems on associative functions. Aequationes Math. 66, 128–140 (2003)
3. Alsina, C., Frank, M.J., Schweizer, B.: Associative Functions: Triangular Norms and Copulas. World Scientific Publishing Co, Singapore (2006)
4. Anderson, T.W., Darling, D.A.: A test of goodness of fit. J. Amer. Statist. Assoc. 49, 765–769 (1954)
5. Blum, J.R., Kiefer, J., Rosenblatt, M.: Distribution-free tests of independence based on the sample distribution function. Ann. Math. Statist. 32, 485–498 (1961)
6. Darsow, W.F., Frank, M.J.: Associative functions and Abel-Schröder systems. Publ. Math. Debrecen. 30, 253–272 (1983)
7. Erdely, A.: Diagonal Properties of the Empirical Copula and Applications. Construction of Families of Copulas with Given Restrictions. Ph.D. Dissertation, Universidad Nacional Autónoma de México (2007)
8. Erdely, A., González-Barrios, J.M.: On the construction of families of absolutely continuous copulas with given restrictions. Comm. Statist. Theory Methods 35, 649–659 (2006)
9. Frank, M.J.: Diagonals of copulas and Schröder's equation. Aequationes Math. 51, 150 (1996)
10. Fredricks, G.A., Nelsen, R.B.: Copulas constructed from diagonal sections. In: Beneš, V., Štěpán, J. (eds.) Distributions with Given Marginals and Moment Problems, pp. 129–136. Kluwer Academic Publishers, Dordrecht (1997)
11. Hollander, M., Wolfe, D.A.: Nonparametric Statistical Methods. Wiley-Interscience, New York (1999)
12. Kallenberg, W.C.M., Ledwina, T.: Data-driven Rank tests for independence. J. Amer. Stat. Assoc. 94, 285–301 (1999)
13. Kolesárová, A., Mesiar, R., Mordelová, J., Sempi, C.: Discrete Copulas. IEEE Trans. Fuzzy Syst. 14, 698–705 (2006)
14. Kolesárová, A., Mordelová, J.: Quasi-copulas and copulas on a discrete scale. Soft Computing 10, 495–501 (2006)

15. Mayor, G., Suñer, J., Torrens, J.: Copula-like operations on finite settings. IEEE Trans. Fuzzy Syst. 13, 468–477 (2005)
16. Mesiar, R.: Discrete copulas - what they are. In: Montseny, E., Sobrevilla, P. (eds.) Joint EUSFLAT-LFA 2005 Conference Proceedings. Universitat Politecnica de Catalunya, Barcelona, pp. 927-930 (2005)
17. Nelsen, R.B.: An Introduction to Copulas, 2nd edn. Springer, New York (2006)
18. Nelsen, R.B., Quesada-Molina, J.J., Rodríguez-Lallena, J.A., Úbeda-Flores, M.: On the construction of copulas and quasi-copulas with given diagonal sections. Insurance Math. Econom. 42, 473–483 (2008)
19. Schweizer, B., Sklar, A.: Probabilistic Metric Spaces. North-Holland, New York (1983)
20. Sklar, A.: Fonctions de répartition à *n* dimensions et leurs marges. Publ. Inst.Statist. Univ. Paris 8, 229–231 (1959)
21. Sungur, E.A., Yang, Y.: Diagonal copulas of Archimedean class. Comm. Statist. Theory Methods 25, 1659–1676 (1996)