Gazi University Journal of Science
GU J Sci
25(2): 417-424 (2012)

ORIGINAL ARTICLE

GUJS

An Application of Archimedean Copulas for Meteorological Data

Vadoud NAJJARI[1,♠], Mehmet Güray ÜNSAL[2]

[1]*Islamic Azad University, Maragheh Branch, Maragheh, Iran*
[2]*Gazi university, Department of Statistics, Ankara, Turkey*

Received: Revised: Accepted:

ABSTRACT

In this study, we show an application of Archimedean copulas for meteorological data, which may be of interest in its own right. Hence we take monthly lowest and highest air temperature records from 1951 to 2005 in Tehran, and then we investigate a suitable family of Archimedean copulas to fit this data. Then we characterize the interval of suitable, and also the best copula parameter. Moreover, basic properties of Archimedean copulas will be presented.

Keywords: Archimedean copulas, Goodness of fit test, Meteorological data.

1. INTRODUCTION

Copulas are a powerful tool in multivariate statistics. If copula functions are used for modeling dependence between random variables, there is an immediate and obvious need to test whether the model can actually describe the data at hand accurately enough or not.

Archimedean copulas are very easy to construct, many parametric families belong to this class and have great variety of different dependence structures, hence many authors are working with these copulas, and also most of them investigate the bivariate case. Emura et al. [8], worked on goodness-of-fit testing procedure for Archimedean copulas (AC) models and they extended the existing method, which is suitable for the Clayton model, and also to the general AC models. Many other authors worked on Archimedean copulas like, Genest and MacKay [11,12], Nelsen [18] etc. There are many references concerning the statistical inference of copulas. In estimation of the right copula, different approaches have been established: Joe and Xu [17] estimate a parametric family by maximum likelihood. For the class of Archimedean copulas, Genest and Rivest [10] developed a semiparametric estimator.

In this study, we use monthly lowest and highest air temperature records from 1951 to 2005 in Tehran-Iran [16] (650 data) and to investigate the right copula for

♠Corresponding author, email: vnajjari@iau-maragheh.ac.ir

modeling of this data, we concentrate on the method by Genest & Rivest [10] and also Çelebioğlu [2].

This paper is structured as follows: Section 2 introduces the family of Archimedean copulas and their properties. In Section 3 we see goodness-of-fit tests for parametric families of Archimedean copulas and determine which copula is suitable to use. In Section 4, some conclusions are given.

2. ARCHIMEDEAN COPULAS – BASIC PROPERTIES, SOME WELL KNOWN FAMILIES

One of the most important classes of copulas is known as Archimedean copulas. There are a number of reasons why these find a wide range of applications in practice. These copulas are very easy to construct and many parametric families belong to this class. Archimedean copulas allow for a great variety of different dependence structures and all commonly encountered Archimedean copulas have simple closed form expressions. In addition, the Archimedean representation allows us to reduce the study of a multivariate copula to a single univariate function. Now basic properties of Archimedean copulas are presented below. More information could be found in Nelsen [18].

Definition 2.1. Let φ be a continuous, strictly decreasing function from [0,1] to $[0,\infty]$ such that $\varphi(1)=0$. The pseudo-inverse of φ is the function $\varphi^{[-1]}$ given by,

$$\varphi^{[-1]}(t)=\begin{cases}\varphi^{(-1)}(t), & 0\le t\le\varphi(0)\\ 0, & \varphi(0)<t\le\infty\end{cases}$$

Definition 2.2. Copulas of the form $C(u,v)=\varphi^{[-1]}(\varphi(u)+\varphi(v))$ for every u,v in [0,1], are called Archimedean copulas, and the function φ is called a generator of the copulas. If $\varphi[0]=\infty$, we say that φ is a strict generator. In this case, $\varphi^{[-1]}=\varphi^{-1}$ and $C(u,v)=\varphi^{-1}\big(\varphi(u)+\varphi(v)\big)$ is said to be a strict Archimedean copula.

Example 2.3. Assume $\varphi(t)=-ln(t)$ for $t\in[0,1]$. Since $\varphi[0]=1$, the inverse is given by $\varphi^{-1}(t)=exp(-t)$ and therefore,

$$C(u,v)=exp(-[(-lnu)+(-lnv)])=uv=\prod(u,v)\ .$$

Thus $\prod$ is an Archimedean copula and $\varphi(t)=-ln(t)$ is its generator.

***Recall the Archimedean axiom for the positive real numbers*:**

If a,b be positive real numbers, then there exists an integer n such that $na<b$. An Archimedean copula behaves like a binary operation on the interval [0,1], in that the copula C assigns to each pair u,v in I a number $C(u,v)$ in I (see [18]). Archimedean copulas feature some algebraic properties which build the content of the next theorem [18].

Theorem 2.4. Let C be an Archimedean copula with generator φ. Then:

1. C is symmetric; i.e., $C(u,v)=C(u,v)$ for all u,v in[0,1].

2. C is associative, i.e., $C(C(u,v),w)=C(u,C(v,w))$ for all u,v,w in [0,1].

3. If $c>0$ is any constant, then $c\varphi$ is also a generator of C.

From the third property we can see that a generator φ uniquely determines an Archimedean copula only up to a scalar multiple. Now we recall some important families of Archimedean copulas, and their graphs, which can be suited to our aim (see Nelsen [18]).

1. Clayton family

This family of copulas firstly was discussed by Clayton [3], after Cox and Oakes [7], also Cook and Johnson [5,6] used it. Genest and MacKay [11,12] call this the generalized Cook and Johnson family; Hutchinson and Lai [15] call it the Pareto family of copulas, while Genest and Rivest [10] call it the Clayton family. This family is as below,

$$C_\theta(u,v)=max\left(\left[u^{-\theta}+v^{-\theta}-1\right]^{\frac{-1}{\theta}},0\right)$$

$$\varphi_\theta=\frac{1}{\theta}\left(t^{-\theta}-1\right)\ ,\ \theta\epsilon[-1,\infty)-\{0\}$$

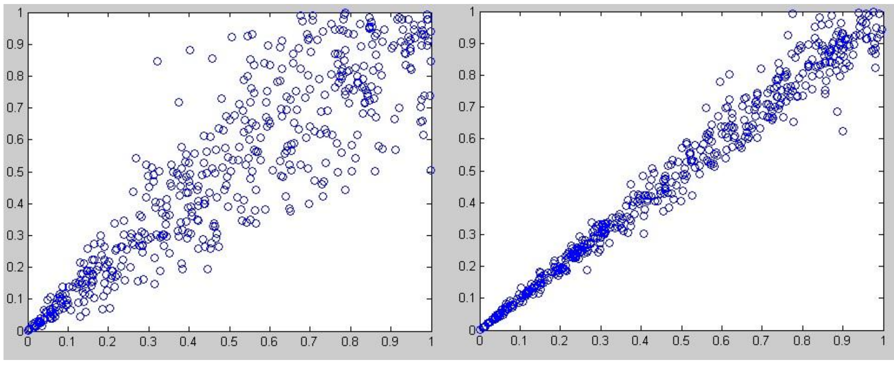

Figure 1. Scatterplots for Clayton copula, $\theta=5$ (left) and $\theta=20$ (right), $n=700$.

*2. **Frank family***

This family first appeared in Frank [9] in a non-statistical context. Some of the statistical properties of this family were discussed in Nelsen [18]. Copulas of this family are as below:

$$C_\theta(u,v) = \frac{1}{\theta} ln\left(1 + \frac{(e^{-\theta u} - 1)(e^{-\theta v} - 1)}{e^{-\theta} - 1}\right),$$

$$\varphi_\theta - ln\left(\frac{e^{-\theta t} - 1}{e^{-\theta} - 1}\right) \ , \ \theta \epsilon (-\infty, \infty) - \{0\}$$

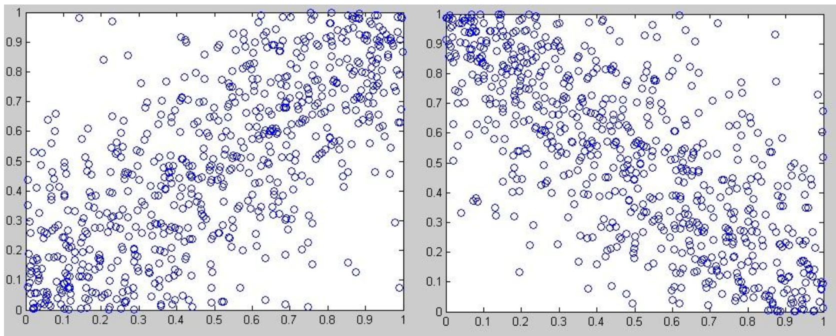

Figure 2. Scatterplots for Frank copula, $\theta = 6$ (left) and $\theta = -6$ (right), $n = 700$.

3. Ali- Mikhail- Haq family

This family is derived algebraically and we can see it in Ali M.M. et. al. [1]. This family of copulas has the functional form and generator, respectively:

$$C_\theta(u,v) = \frac{uv}{1 - \theta(1-u)(1-v)}$$

$$\varphi_\theta = ln\frac{1 - \theta(1-t)}{t} \ , \ \theta \epsilon [-1,1)$$

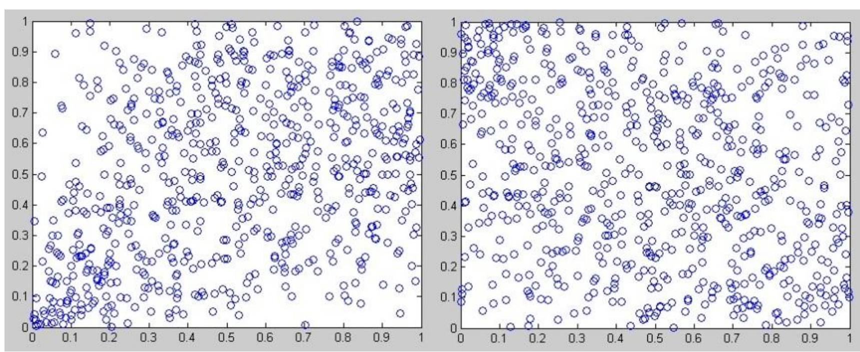

Figure 3. Scatterplots for Ali-M-Haq family, $\theta = 1$ (left) and $\theta = -1$ (right), $n = 700$.

4. Copulas family of 4.2.12

This family hadn't name and in Nelsen [18] is numbered as 4.2.12. Copulas of this family are as the following:

$$C_\theta(u,v) = \left(1 + \left[(u^{-1} - 1)^\theta \ + \ (v^{-1} - 1)^\theta\right]^{\frac{1}{\theta}}\right)^{-1}$$

$$\varphi_\theta = \left(\frac{1}{t} - 1\right)^\theta \ , \ \theta \epsilon [1, \infty).$$

See Figure 4.

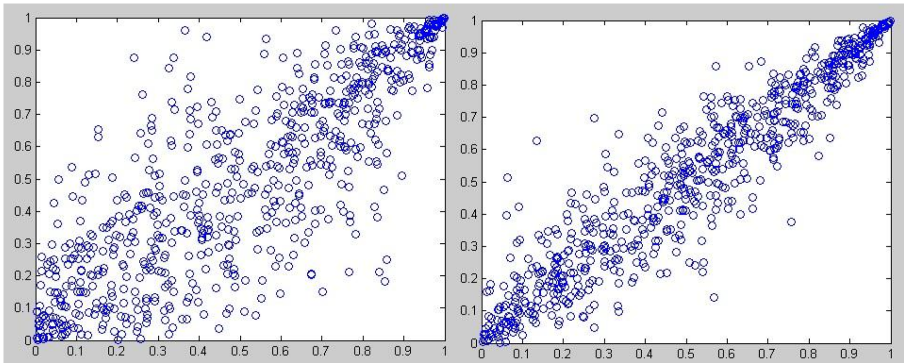

Figure 4. Scatterplots for the family 4.2.12 , $\theta = 1.5$ (left) and $\theta = 4$ (right), $n = 700$.

5. Copulas family of 4.2.16

This family hadn't name and in Nelsen [18] and is numbered as 4.2.16. Copulas of this family are as below:

$$C_\theta(u,v) = \frac{1}{2}\left(S + \sqrt{S^2 + 4\theta}\right),$$

$$S = u + v - 1 - \theta\left(\frac{1}{u} + \frac{1}{v} - 1\right),$$

and $\varphi_\theta = \left(\frac{\theta}{t} + 1\right) \quad (1 - t)$, $\theta \epsilon [0, \infty)$

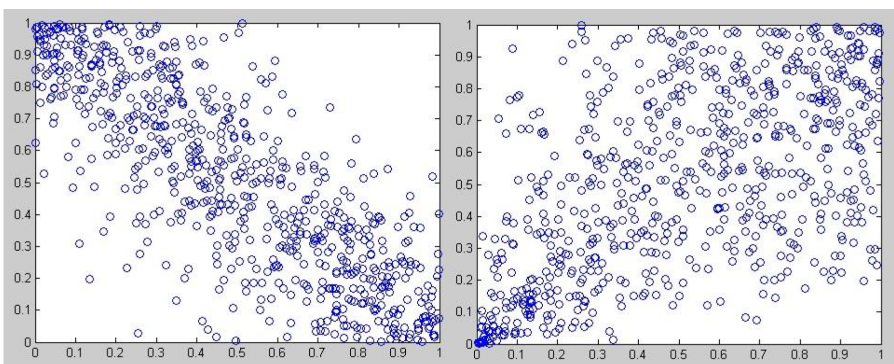

Figure 5. Scatterplots for the family 4.2.16, $\theta = 0.01$ (left) and $\theta = 1$ (right), $n = 700$.

6. Plackett family

Plackett family [19] copulas have been widely used both in modeling and as alternatives to the bivariate normal for studies of power and robustness of various statistical tests (Conway [4], Hutchinson and Lai [15]). Copulas of this family are as below (see Figure 6):

$$C_\theta(u,v) = \frac{\left(1 + (\theta - 1)(u + v)\right) - \sqrt{\left(1 + (\theta - 1)(u + v)\right)^2 - 4uv\theta(\theta - 1)}}{2(\theta - 1)}, \qquad \theta \epsilon [0, \infty)$$

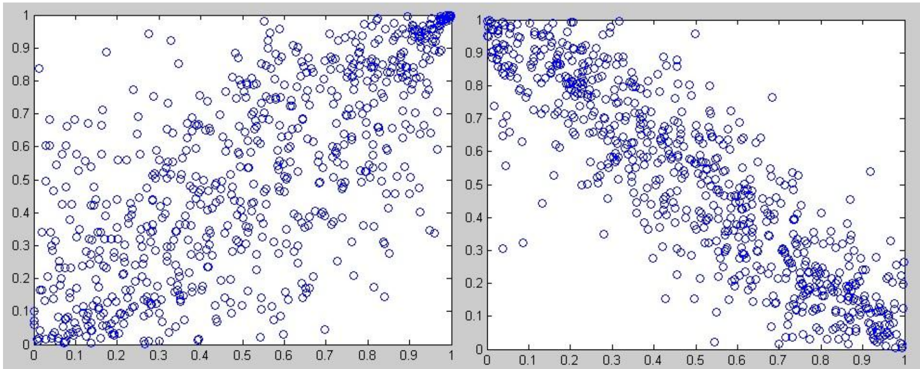

Figure 6. Scatterplots for Plackett family, $\theta = 20$ (left) and $\theta = 0.02$ (right), $n = 700$.

7. Gumbel's family

This family of copulas was first discussed by Gumbel [13], hence many authors refer to it as the Gumbel's family. However, because Gumbel's name is attached to another Archimedean family (4.2.9 in Nelsen's book [18]), this family also appears in Hougaard [14], Hutchinson and Lai [15] refer to it as the Gumbel-Hougaard family. Copulas of this family are as the following:

$$C_\theta(u,v) = exp\left(-[(-lnu)^\theta + (-lnv)^\theta]^{\frac{1}{\theta}}\right), \qquad \varphi_\theta = (-lnt)^\theta \ , \ \theta\epsilon[1,\infty)$$

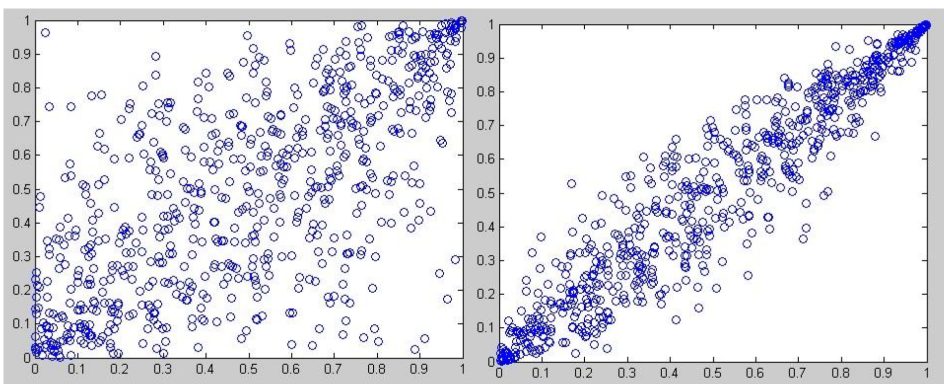

Figure 7. Scatterplots for Gumbel family, $\theta = 2$ (left) and $\theta = 4$ (right), $n = 700$.

3. DATA DESCRIPTION AND RESULTS

To find a relationship in the series of our data (monthly lowest and highest air temperature, on the other hand, finding (u_i, v_i), in each series, first any two ordered data has been subtracted, then the new data is arranged, finally it has been divided to total data sample size plus one, as below (see [2,10]):

$$u_i = \frac{R(x_i)}{n+1} \ , \ v_i = \frac{R(y_i)}{n+1}, \quad i = 1,2,\dots,n$$

where x_i is the lowest, y_i is the highest air temperature records and $R(x_i), R(y_i)$ are the rank (in ascending order) of data. Our resulted data (u and v) will be in interval (0,1). Figure 8 shows scatterplots of the result data.

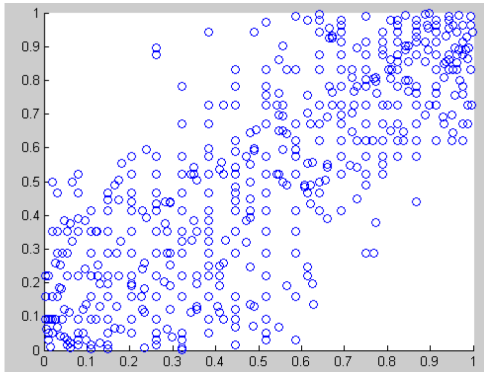

Figure 8. Scatterplots of air temperature data

By using relation $\sqrt[4]{n}$, (n number of data), (see Çelebioğlu [2]), we divided the range of two variables uniform transformation into 5 intervals each, therefore $df = 16$, and $\chi^2_{0.05,\ 16} = 26.2962$, thus copulas family will be suit if, $\chi^2 < \chi^2_{0.05\ ,16}$ (where $\chi^2 = \sum \frac{(O_{i,j}-E_{i,j})^2}{E_{i,j}}$). To find a suitable family of Archimedean copulas with special parameter, for the all seven copula families mentioned earlier, Matlab software has been used and we calculated χ^2 test statistic value, which the results are summarized in Table1. Moreover, we will separately discuss about the selected family.

- Frank family

As we can see in Figure 9 and Table 1, for θ values less than 7.7 the related function is decreasing and for more than 7.7, it is increasing and minimum value is in the point θ = 7.7. Thus, this family is suitable with $6.9 < \theta < 8.6$ and the best value for θ is 7.7, which value of the related χ^2 is 23.781. For this point we see observed and expected matrix as below,

$$O_{i,j} = \begin{pmatrix} 63 & 40 & 14 & 1 & 0 \\ 41 & 55 & 40 & 7 & 3 \\ 24 & 25 & 42 & 26 & 12 \\ 2 & 6 & 32 & 49 & 38 \\ 0 & 1 & 5 & 49 & 75 \end{pmatrix} \qquad E_{i,j} = \begin{pmatrix} 80.8709 & 35.7441 & 10.1337 & 2.3409 & 0.5104 \\ 35.7441 & 50.8523 & 30.8477 & 9.8150 & 5.3409 \\ 10.1337 & 30.8477 & 47.6373 & 30.8477 & 10.1337 \\ 2.3409 & 9.8150 & 30.8477 & 50.8523 & 35.7441 \\ 0.5104 & 5.3409 & 10.1337 & 35.7441 & 80.8709 \end{pmatrix}$$

4. CONCLUSION

For monthly the lowest and highest air temperature records from 1951 to 2005 in Tehran, we investigated a suitable family of Archimedean copulas to fit this data. Hence we analyzed seven well-known families of Archimedean copulas and we saw that, only Frank family is suitable for this data with the best parameter value 7.7 with the related $\chi^2 = 23.781$.

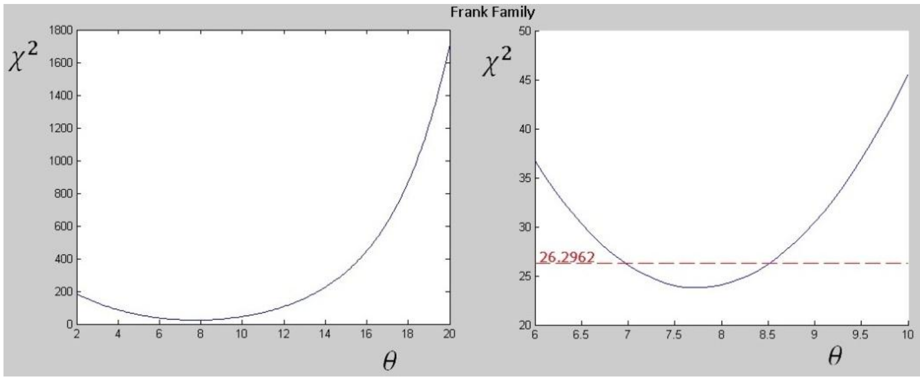

Figure 9. Graph of (θ, χ^2) in Frank family.

Table 1. Calculated χ^2 for all seven copula families.

Frank		Gumbel		Ali. M. hag		Plackett		4.2.16		4.2.12		Clayton	
θ	χ^2	θ	χ^2	θ	χ^2	θ	χ^2	θ	χ^2	θ	χ^2	θ	χ^2
1	255.880	1	203.785	-1	915.056	1	0.000	1	600.245	1	106.360	0.2	179.623
3	128.194	1.5	103.985	-0.8	744.074	1.1	57944.043	4	265.607	1.1	94.432	0.8	121.199
5	56.641	2	57.374	-0.6	623.661	1.5	1566.665	7	122.116	1.2	97.222	1	106.360
6	36.750	2.5	40.934	-0.5	574.336	1.9	205.952	8	108.168	1.3	111.567	1.6	75.273
6.9	26.761	2.6	40.261	-0.2	452.150	2	148.240	8.2	106.815	1.4	136.111	1.8	69.155
7	26.090	2.7	40.331	-0.1	417.272	2.1	133.052	8.4	105.886	1.6	216.021	2	65.063
7.5	24.017	3	44.650	0	384.431	2.2	149.210	8.6	105.361	1.8	345.635	2.2	62.969
7.6	23.857	5	223.936	0.3	294.967	2.5	326.668	8.8	105.222	2	546.244	2.4	62.874
7.7	23.781	7	959.700	0.4	267.447	3	902.055	9	105.448	3	4965.438	2.6	64.810
7.8	23.788	10	9215.641	0.5	240.871	5	5470.200	9.2	106.023	5	390396.761	3	75.090
8	24.056	12	46305.819	0.6	215.240	8	18280.262	10	111.491	7	26485572.225	4	146.617
8.5	26.195	14	244876.583	0.7	190.715	10	30718.175	13	165.622	8	207178333.259	6	712.477
8.6	26.875	15	570228.547	0.8	167.800	15	75427.000	14	191.984	9	1579662268.278	8	3404.989
9	30.453	18	7422193.518	0.9	147.855	20	139580.000	15	221.274	10	11810372190.775	9	7420.560
15	319.558	20	41652142.495	1	134.857	30	326200.000	20	397.281	15	249410000000000	10	16142.120

REFERENCES

[1] Ali, MM,. Mikhail, NN,.Haq, MS,. A class of bivariate distributions including the bivariate logistic. Multivariate Anal. 8: 405-412 (1978).

[2] Çelebioğlu, S,. Archimedean Copulas And An Application, S Ü Fen Ed Fak Fen Derg , 22 (2003) 43- 52, KONYA.

[3] Clayton, DG,. A model for association in bivariate life tables and its applicationin epidemiological studies of familial tendency in chronic disease incidence. Biometrika, 65: 141-151(1978).

[4] Conway, DA,. Plackett family of distributions. In: Kotz S. Johnson NL (eds). Encyclopedia of Statistical Sciences, Vol 7. Wiley, New York, 1-5 (1986).

[5] Cook RD, Johnson ME,. A family of distributions for modeling non-elliptically symmetric multivariate data. J. Roy Statist. Soc. SerB 43:210-218(1981).

[6] Cook RD,. Johnson ME,.Generalized Burr-Pareto-logistic distributions withapplications to a uranium exploration data set. Technometrics 28:123-131(1986).

[7] Cox DR,. Oakes D,. Analysis of Survival Data. Chapman and Hall, London (1984).

[8] Emura T,. LinCW,. Wang W,. A goodness-of-fit test for Archimedean copula models in the presence ofright censoring. Computational Statistics and Data Analysis, (2008).

[9] Frank MJ,. On the simultaneous associativity of $F(x, y)$ and $x + y - F(x, y)$. Aequationes Math 19:194-226 (1979).

[10] Genest,C., Rivest L.P,. Statistical Inference Procedures for Bivariate Archimedean Copulas. Journal of the American Statistical Association, Vol. 88, No. 423, pp. 1034-1043 (1993).

[11] Genest C,.MacKay J,. Copules archimédien nesetfamilles de loisbidimension nellesdont les margessontdonnées. Canad. J. Statist., 14:145-159 (1986a).

[12] Genest C,. MacKay J,. The joy of copulas: Bivariate distributions with uniform marginals. Amer. Statist. 40:280-285 (1986b).

[13] Gumbel EJ,. Distributions des valeursextrêmes en plusiers dimensions. PublInst. Statist. Univ. Paris 9:171-173 (1960b).

[14] Hougaard P,. A class of multivariate failure time distributions. Biometrika, 73:671-678 (1986).

[15] Hutchinson TP,. Lai CD,. Continuous Bivariate Distributions, Emphasising Applications. Rumsby Scientific Publishing, Adelaide (1990).

[16] I.R. OF Iran meteorological organization (IRIMO).

[17] Joe, H,. Multivariate Models and Dependence Concepts. Chapman & Hall, London (1997).

[18] Nelsen, R.B,. An Introduction to Copulas, Springer, New York (1999).

[19] Plackett RL,. A class of bivariate distributions. J Amer. Statist. Assoc, 60:516-522(1965).