Archives of Current Research International

21(2): 14-21, 2021; Article no.ACRI.66579
ISSN: 2454-7077

A Study on Modelling of Bivariate Competing Risks with Archimedean Copulas

Cigdem Topcu Guloksuz[1*]

[1]*Department of Management Information Systems, Ankara Science University, 06200 Ankara, Turkey.*

Author's contribution

The sole author designed, analyzed, interpreted and prepared the manuscript.

Article Information

DOI: 10.9734/ACRI/2021/v21i230229
Editor(s):
(1) Dr. Wang Mingyu, Central South University, China.
Reviewers:
(1) Róbert Csalódi, University of Pannonia, Hungary.
(2) Ahlam H. Tolba, Mansoura University, Egypt.
(3) Elkhan Richard Sadik-Zada, Ruhr-University Bochum, Germany.
(4) Hugo Cruz-Suárez, Benemérita Universidad Autónoma de Puebla, Mexico.
Complete Peer review History: http://www.sdiarticle4.com/review-history/66579

Original Research Article

Received 11 February 2021
Accepted 22 April 2021
Published 29 April 2021

ABSTRACT

In this study we consider Archimedean copula functions to obtain estimates of cause-specific distribution functions in bivariate competing risks set up. We assume that two failure times of the same group are dependent and this dependency can be modeled by an Archimedean copula. Based on the Archimedean copula which gives best fit to the competing risk data with independent censoring we obtain the estimates of cause specific sub distributions.

Keywords: Archimedean copulas; cause specific distribution; cumulative incidence function; competing risks; nonparametric estimation.

1. INTRODUCTION

In the classical competing risk setting, there are two observed outcomes which one is time to failure ($T > 0$) and the other one is cause of failure ($C = \{1,2, \dots, k\}$). Here, T is considered as a continuous variable and C is considered as a discrete variable. Under the assumption of that every failure is assigned to only one cause, the set of k causes are called *risks* before the failure occurs and then they are called *causes* after the failure occurs. In other words, the risks compete to be the cause [1].

The basic mathematical framework of bivariate competing risk setting can be defined as a joint

**Corresponding author: Email: cigdem.topcu.guloksuz@ankarabilim.edu.tr;*

distribution of (T, C) . There are two basic quantities to analyze competing risks data which are the cause specific hazard $(\lambda_j(t))$ and the cause specific sub distribution or cumulative incidence function (CIF), $(F_j(t))$. These quantities are defined for the cause j by time t , respectively.

$$\lambda_j(t) = \lim_{\Delta t \to 0} \frac{P(T < t + \Delta t, C = j | T \geq t)}{\Delta t};\ j = 1,2, \dots, k \quad (1)$$

$$F_j(t) = P(T \leq t | C = j);\ j = 1,2, \dots, k \quad (2)$$

Competing risks data commonly occur in biomedical, epidemiological, medical studies and analyzing such kind of data becomes an important task for the goal of the researches. One of the procedures of analyzing competing risks data is estimating CIF and consequently compare estimated CIFs. Therefore, Kaplan-Meier method which is a well known nonparametric method in survival analysis to estimate survival function from lifetime data is also considered to obtain estimates of CIFs in competing risks data. The cases with no competing risks, one minus Kaplan-Meier estimates of survival function provides an estimate of CIF. However, in the cases with competing risks Kaplan-Meier method to estimate CIFs yields incorrect and unbiased results. Because, it is assumed in the traditional approach in competing risk model all causes of failure and consequently failure times are considered independent. It has been studied by [2-4]. We recommend [1,5-7] for comprehensive sources for analysis of competing risks data.

The competing risks models can be defined via latent failure times representation. In the presence of competing risks, the basic assumption is that there is a potential $T = (T_1, T_2, \dots, T_k)$ survival time vector which is assigned to every individual. Since these failure times are not observable they have a latent structure and the actual lifetime span is the minimum of the $, T_2, \dots, T_k$. Mathematically, these failure times are treated as non-negative random variables and their joint distribution function is a multivariate distribution. Many authors state that in real life applications $T_1, T_2, \dots, T_k$ are likely to be dependent and this dependency can be modelled by copula functions, [8-12].

In some studies, multivariate competing risks data with multiple cause of failure can be obtained from the subjects in the same group or family. Estimating the joint distribution of such kind of multivariate data provides the better understanding of the dependence among failure times.

In this study, we consider bivariate competing risks set up in which dependence structure of failure times can be modelled by an Archimedean copula. The Archimedean copula function that provides best fit for the dependency between two dependent competing risks is estimated non-parametrically, by the method that is suggested by [13].

Under independent censoring a non-parametric estimation of cause-specific sub distribution which is developed in [14] is employed and the same data set is used for comparison of the empirical results.

In this study, we revisited the approach in the work of [12] in which Archimedean copulas are considered to model failure times in the presence of competing risks. Following the approach we model the dependency of the competing risks by an Archimedean copula and obtain the estimates of cause-specific sub distributions. Under independent censoring and Archimedean copula approach the obtained estimates are compared to the estimates in [14].

The rest of the paper is organised as follows. In Section 2, a brief mathematical framework of the bivariate competing risks and Archimedean copulas are revisited. In Section 3, the results are obtained and compared. In Section 4 we conclude the paper.

2. THE MATHEMATICAL FRAMEWORK

2.1 Bivariate Competing Risks Models

The basic probability framework of bivariate competing risks is considered as a bivariate distribution. Let, T_1 and T_2 are failure times of each member of the same group and random variables on a probability space. The survivor function of $T = (T_1, T_2)$ is defined as

$$S_T(t_1, t_2) = P(T_1 > t_1, T_2 > t_2),\ 0 < t_1, t_2 < \infty \quad (3)$$

From the nature of the failure time data we can observe the $Y_i = min(T_i, Z_i)$ where $\delta_i = I(T_i = Y_i)$ with pair of random censoring times, $Z = (Z_1, Z_2)$. The assumption here is T and Z are independent. The survivor function of $Y = (Y_1, Y_2)$ is

$$S_Y(t_1,t_2) = P(Y_1 > t_1, Y_2 > t_2);\ 0 < t_1, t_2 < \infty \quad (4)$$

In bivariate competing risk set up, suppose that the causes $C = (C_1, C_2)$ corresponding to $T = (T_1, T_2)$ are represented. The cause specific sub distribution function is defined as follows

$$F_{ij}(t_1, t_2) = P(T_1 \le t_1, T_2 \le t_2, C_1 = i, C_2 = j); \\ i = 1,2,\dots,k;\ j = 1,2,\dots,k \quad (5)$$

The cause specific sub distribution function can be interpreted as the probability that the failure of both the individuals due to the causes (i,j) is prior (t_1, t_2). Under the independent censoring, the cause specific sub distribution function, F_{ij} is defined in (6).

$$F_{ij}(t_1,t_2) = P(T_1 \le t_1, T_2 \le t_2, \delta_1 = 1, \delta_2 = 1, C_1 = i, C_2 = j) \quad (6)$$

The simple non parametric estimator of (6) is defined and presented with the properties in [14]. We consider the following unbiased estimator rest of the paper. Suppose that random sample with size of n and $Y_u = (Y_{1u}, Y_{2u})$, $\delta_u = (\delta_{1u}, \delta_{2u})$ with cause of failure pair $C_u = (C_{1u}, C_{2u})$. The unbiased estimator of F_{ij} is defined by (7).

$$\hat{F}_{ij} = \frac{1}{n}\sum_{u=1}^{n} I(Y_{1u} \le t_1, Y_{2u} \le t_2, \delta_{1u} = 1, \delta_{2u} = 1, C_{1u} = i, C_{2u} = j) \quad (7)$$

2.2 Archimedean Copulas

According to [15], the bivariate cumulative distribution function H of any pair (X,Y) of continuous random variables may be written in the form

$$H(x,y) = C(F(x), G(y)), x, y \in \mathbb{R} \quad (8)$$

where

$F(x)$ and $G(y)$ are continuous marginal distributions and C is the copula function with $C: [0,1]^2 \to [0,1]$. It should be noted that if the marginal distributions are continuous, there is a unique copula representation [15].

Archimedean Copula family which is a special class of copulas can be expressed in the following form

$$C(x,y) = \phi^{-1}\big(\phi(F(x)) + \phi(G(y))\big) \quad (9)$$

Here, $\phi: [0,1] \to [0,\infty]$ is called a generator function which is a convex, decreasing function. For $0 < t < 1$, the generator function ϕ is such that $\phi(1) = 0$, $\phi(0) = \infty$, $\phi'(\mathrm{t}) < 0$, $\phi''(\mathrm{t}) \ge 0$. Thus, ϕ is a continuous, strictly decreasing and convex function and always has an inverse, $\phi^{-1}: [0,\infty] \to [0,1]$ which has the same properties except $\phi^{-1}(0) = 1$ ve $\phi^{-1}(\infty) = 0$.

An Archimedean copula is indexed by a parameter (θ) which is called copula parameter. The copula parameter can be estimated by using some estimation methods such as maximum likelihood and method of moments which is based on Kendall's tau. The performances of these two methods are compared and it is stated that the method of moments based on Kendall's tau performs as well as the maximum likelihood method does, [12].

Additionally, the direct relationship between Kendall's tau and Archimedean copulas provides some benefits for mathematical calculations. The relationship between parameters of some Archimedean copulas and Kendall's tau is defined in [16].

$$\tau = 4\iint_0^1 C(u,v)\,dC(u,v) - 1 \quad (10)$$

We consider four Archimedean copula functions [5,17-19] to model failure times. These copula functions, corresponding parameter space and the relationship with Kendall's tau are listed in Table 1.

Table 1. Archimedean copulas and Kendall's τ.

Family	Bivariate copula $C(u_1,u_2)$	Copula Parameter Space	τ
Gumbel	$exp\left[-[(-lnu_1)^\theta + (-lnu_2)^\theta\ ]^{1/\theta}\right]$	$\theta \ge 1$	$(\theta-1)/\theta$
Clayton	$\left((u_1)^{-\theta} + (u_2)^{-\theta} - 1\right)^{-1/\theta}$	$\theta > 1$	$\theta/(\theta+2)$
Frank	$\left(-\frac{1}{\theta}\right) ln\left\{\frac{(1-e^{-\theta}) - (1-e^{-\theta u_1})(1-e^{-\theta u_2})}{(1-e^{-\theta})}\right\}$	$\theta \epsilon (-\infty, \infty)$	$1 - \frac{4}{\theta}[1 - D_1(\theta)]$
Guloksuz-Kumar	$1 - \left(\frac{1}{\theta}\right) ln\left[e^{\theta(1-u_1)} + e^{\theta(1-u_2)} - 1\right]$	$\theta > 0$	$4\left(\frac{1-e^{-\theta}-\theta}{\theta^2}\right) + 1$

$D_n(\Theta) = \frac{n}{\Theta^n}\int_0^{\Theta} \frac{t^n}{e^{t-1}}dt, n > 0$, *is a Debye function*

A bivariate Archimedean copula that is defined in (9) can be uniquely determined by Kendall distribution function, $K_\phi(t)$, [13] as follows :

$$K_\phi(t) = t - \frac{\phi(t)}{\phi'(t)}. \tag{11}$$

It means that a bivariate Archimedean copula function can be determined by one-dimensional (t) , [13]. In fact, $K_\phi(t)$ is the distribution function of the Archimedean copula function. The considered copula functions in this study and their distribution functions are listed in Table 2.

Under the assumption of two failure times can be modelled by a bivariate Archimedean copula function, we consider the method which is proposed in [13] to select suitable Archimedean copula. The method is based on the comparison of empirical and theoretical estimates of Kendall distribution in (11) which uniquely determine a bivariate Archimedean copula function.

As stated in [13], the empirical estimate of (11) which is represented by $K_n(t)$ from a random sample of size *n* is given by

$$K_n(t) = \frac{\#(T_i \le t)}{n+1} \tag{12}$$

where pseudo observations $T_i's$ are defined as

$$T_i = H_n(X_i, Y_i) = \frac{\sum_{j=1}^{n} I[\{X_j \le X_i \& Y_j \le Y_i\}]}{n+1}, i = 1,2, \dots n \tag{13}$$

The theoretical estimates of (11) are obtained by considering Table 1 which provides the estimator of copula parameter and third column of Table 2. The selection procedures can be finalized, following [20], minimizing the following distance which specifies the degree of closeness of the $K_n(t)$ and $K_\phi(t)$ in this study.

$$MD = \int [K_n(t) - K_\phi(t)]^2 dK_n(t) \tag{14}$$

3. APPLICATION

In this section we consider the data set which refers the times to tumor appearance or death for 50 pairs of mice from the same litter in a tumor genesis experiment, [21], as reported in [22]. In this data set, T_1 and T_2 indicate the failure times (in weeks) of mice , $C_j = 1$ represents the cause of failure is the appearance of a tumour , $C_j = 2$ represents that the failure is observed before the appearance of a tumour and $C_j = 0$ represents the censored observations. Since the length of the observation period is 104 weeks, censoring time is 104 weeks for all mice. The all records of the data set are listed in Table 3. The table also can be found in [14].

We assume that the dependence structure of the failure times can be modelled by an Archimedean copula. The considered copula functions in this study are listed in Table 1 and Table 2. The procedure which is summarized briefly in Section 2.2 is applied to estimate the bivariate Archimedean copula which gives the best fit to the data. The estimates of the parameters of the studied copulas and the distance measure in (14) are listed in Table 4.

According to the results, Frank copula with the estimated parameter $\hat{\boldsymbol{\theta}} = 0.7592$ gives the best fit to the data over the considered candidates and Fig. 1 demonstrates the comparison of empirical and theoretical estimates.

In the rest of this section, estimates of the cause specific distribution functions of the studied data are presented. Our aim is comparing the estimates of cause specific distribution functions which are obtained based on (7) and the ones which are obtained with a bivariate Archimedean copula approach. The estimator of cause specific distribution function in (7) is and its mathematical properties are studied in [14]. According to the

Table 2. Archimedean copulas, generator functions and Copula distribution function

Family	Generator $\boldsymbol{\phi(t)}$	$\boldsymbol{K_\phi(t) = t - \frac{\phi(t)}{\phi'(t)}}$
Gumbel	$(-lnt)^\theta$	$t - \frac{tlnt}{\theta}$
Clayton	$t^{-\theta} - 1$	$t - \frac{t^{\theta+1} - t}{\theta}$
Frank	$-ln\frac{e^{-\theta t} - 1}{e^{-\theta} - 1}$	$t - \frac{ln\frac{e^{-\theta t} - 1}{e^{-\theta} - 1}}{\theta}(e^{\theta t} - 1)$

Guloksuz-Kumar	$e^{\theta(1-t)} - 1$	$t - \frac{e^{\theta(1-t)} - 1}{-\theta e^{-\theta(1-t)}}$

Table 3. Data of mice

T1	C1	T2	C2	T1	C1	T2	C2
49	1	104*	0	104*	0	104*	0
102	2	104*	0	104*	0	104*	0
104*	0	104*	0	81	1	64	1
97	2	79	2	55	1	94	2
104*	0	104*	0	104*	0	54	1
96	1	104*	0	87	2	74	2
94	2	77	1	73	1	84	1
104*	0	104*	0	104*	0	83	2
77	2	104*	0	104*	0	73	2
104*	0	77	2	79	2	104*	0
91	2	90	2	104*	0	104*	0
70	2	92	2	104*	0	104*	0
45	2	50	1	101	1	94	2
69	2	91	2	84	1	78	1
104*	0	103	2	81	1	76	2
72	2	104*	0	95	2	104*	0
63	2	104*	0	104*	0	66	1
104*	0	74	2	104*	0	102	1
104*	0	69	2	98	2	73	2
104*	0	68	1	104*	0	104*	0
104*	0	104*	0	83	2	77	2
104*	0	104*	0	104*	0	104*	0
83	2	40	1	79	2	99	2
104*	0	104*	0	91	2	104*	0
104*	0	104*	0	104*	0	79	1

**Censored Observation*

Table 4. Estimates of parameters and distance

Archimedean Copulas	$\hat{\theta}$	Distance
Clayton	0.1831	0.1302
Gumbel	1.0915	0.1118
Frank	0.7592	0.1108
Guloksuz-Kumar	1.6062	1.1067

results which are proposed in [14], the estimator is unbiased, consistent and it has a weak convergence. The simulation studies which are conducted in [15] show that the estimator has low variance and bias.

The results are listed in Table 5 are obtained based on (7). They are also found in [14].

For the aim of the study, the failure times of mice (T_1 and T_2) are modelled by bivariate Frank copula with different parameters and the empirical estimates of the cause specific distribution functions $\hat{F}_{ij}(t_1, t_2)$ are obtained. The results are illustrated in Fig. 2 and listed in Table 6.

The first column of the Table 6 is the estimates of cause specific sub distribution function which are also listed in Table 5. These estimates are obtained based on (7) and the other columns are the estimates under the assumption of the pair of (T_1, T_2) can be modelled by Frank copula with different parameters. We compared the Frank copula based estimates of $\hat{F}_{ij}(t_1, t_2)$ to the estimates which are presented in Table 5 based on (7). The results present that copula based estimates are close to the non-parametric estimates according to the closeness measure Mean Square Error (MSE). It can be said that modelling dependence of competing risks by an Archimedean copula yields reasonable estimates of cause specific distributions.

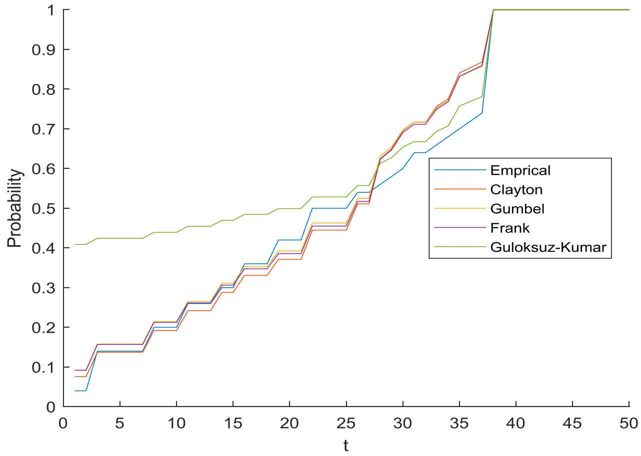

Fig. 1. Estimation of distribution functions of studied copulas and empirical distribution function of data

Table 5. Estimates of the cause specific distribution function

T_1	C_1	T_2	C_2	$\widehat{F}(T_1, T_2)$
45	2	50	1	0.02
55	1	94	2	0.04
69	2	91	2	0.04
70	2	92	2	0.06
73	1	84	1	0.04
79	2	99	2	0.12
81	1	64	1	0.04
81	1	76	2	0.06
83	2	40	1	0.02
83	2	77	2	0.10
84	1	78	1	0.12
87	2	74	2	0.08
91	2	90	2	0.18
94	2	77	1	0.14
97	2	79	2	0.18
98	2	73	2	0.08
101	1	99	2	0.34

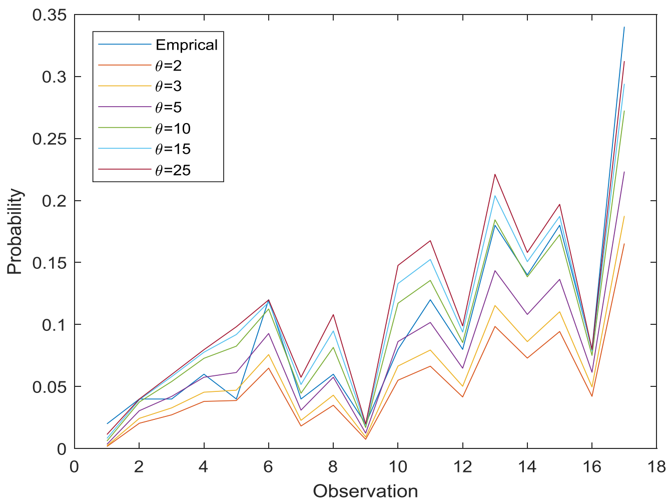

Fig. 2. Estimates of $F_{ij}(t_1, t_2)$ based on frank copula with different parameters

Table 6. Estimates under Frank Copula modelling with different parameters

$\hat{F}(T_1, T_2)$	$\hat{F}_{\hat{\theta}=2}$	$\hat{F}_{\hat{\theta}=3}$	$\hat{F}_{\hat{\theta}=5}$	$\hat{F}_{\hat{\theta}=10}$	$\hat{F}_{\hat{\theta}=15}$	$\hat{F}_{\hat{\theta}=25}$
0,02	0	0	0	0,01	0,01	0,01
0,04	0,02	0,02	0,03	0,04	0,05	0,04
0,04	0,03	0,03	0,04	0,05	0,06	0,06
0,06	0,04	0,05	0,06	0,06	0,08	0,08
0,04	0,04	0,05	0,06	0,05	0,09	0,1
0,12	0,07	0,08	0,09	0,11	0,13	0,12
0,04	0,02	0,02	0,03	0,04	0,05	0,06
0,06	0,04	0,04	0,06	0,07	0,1	0,11
0,02	0,01	0,01	0,01	0,02	0,02	0,02
0,08	0,06	0,07	0,09	0,09	0,13	0,15
0,12	0,07	0,08	0,1	0,12	0,14	0,17
0,08	0,04	0,05	0,07	0,09	0,09	0,1
0,18	0,1	0,12	0,14	0,18	0,19	0,22
0,14	0,07	0,09	0,11	0,14	0,15	0,16
0,18	0,09	0,11	0,14	0,17	0,19	0,2
0,08	0,04	0,05	0,06	0,08	0,08	0,09
0,34	0,16	0,19	0,22	0,32	0,32	0,33
MSE	0,0037	0,0024	0,0012	0,00007	0,0005	0,001

4. CONCLUSION

Most researches in competing risks, failure times are assumed to be independent. However, in some cases they are tend to be dependent and such cases require different approaches. Archimedean copulas can be considered to model the dependency between failure times. In this paper, Archimedean copula functions in bivariate competing risk set up is considered and dependent failure times are modelled by a bivariate Archimedean copula. A comparison study is conducted to present that closeness of estimations of cause specific distribution functions based on Copula and empirical estimates. The empirical estimates of cause specific distributions which are presented in [14] are treated as standard estimates and the copula based estimates are compared to these standard estimates. The results show that when the dependence of failure times is modeled by an Archimedean copula, the estimates of cause specific distribution function are estimated as close as to the standard empirical estimates. The all results are obtained by using one data set. The idea of the study can be improved by considering a comprehensive theoretical perspective. The study can also be extended multivariate case with different estimation methods and copula functions.

ACKNOWLEDGEMENT

The author would like to thank the Chief Editor of the journal and the anonymous reviewers for their constructive comments and suggestions.

COMPETING INTERESTS

Author has declared that no competing interests exist.

REFERENCES

1. Crowder M. Classical Competing Risks, Chapmann & Hall; 2001.
2. Gooley TA, Leisenring W, Crowley J, Storer BE. Estimation of failure probabilities in the resence of competing risks: New representations of old estimators. Statistics in Medicine. 1999;18:695–706.
3. Kim HT. Cumulative incidence in competing risks data and competing risks regression analysis. Clinical Cancer Research. 2007;(13):2.
4. Pepe MS, Modri M. Kaplan-Meier, marginal or conditional probability curves in summarizing competing risks failure time data. Statistics in Medicine. 1993;12:737-751.
5. Clayton DG. A Model for association in bivariate life tables and its application in epidemiological studies of familial tendency in chronic disease

incidence. Biometrika. 1978;65:141–151.
6. Kalbfleisch JD, Prentice RL. The Statistical Analysis of Failure Time Data, Wiley, New York; 1980.
7. Lawless J.F. Statistical Models and Methods for Lifetime Data, 2nd ed, Wiley; 2003.
8. Carriere J. Dependent decrement theory. Transactions, Society of Actuaries XLVI. 1994;45-65.
9. Elandt-Johnson RC. Conditional failure time distributions under competing risk theory with dependent failure times and proportional hazard rates. Scandinavian Actuarial Journal. 1976;37-51.
10. Escarela G, Carriere J. Fitting competing risks with an assumed copula. Statistical Methods in Medical Research. 2003;12(4):333-349.
11. Kaishev KV, Dimitrova DS, Haberman S. Modelling the joint distribution of competing risks survival times using copula functions. Insurance: Mathematics and Economics. 2007;41:339-361.
12. Topçu Ç. Yarışan bağımlı risklerle sağkalım analizinde Archimedean Kapula Yaklaşımı. Ph.D. dissertation, Turkey: Ankara University, Turkish; 2013.
13. Genest C, Rivest LP. Statistical inference procedures for bivariate Archimedean copulas. Journal of The American Statistical Association Theory and Methods 1993;(88):423.
14. Sankaran PG, Lawlees JF, Abraham B, Antony AA. Estimation of distribution function in bivariate competing risk models. Biometrical Journal. 2006;(48):399-410.
15. Sklar A. Functions de repartition a n dimensions et leurs marges. Publ. Inst. Stat. Univ. Paris, French.1959;8:229-231.
16. Nelsen R. An Introduction to copulas. Springer, New York; 1999.
17. Frank MJ. On the simultaneous associativity of F(x,y) and x+y-F(x,y). Aequationes Mathematicae. 197;19:194–226.
18. Guloksuz CT, Kumar P. On a new bivariate one parameter archimedean copula function and ıts application to modeling dependence in climate and life sciences. Statistics and Applications. 2019;17(2): 23-36.
19. Gumbel EJ. Bivariate exponential distributions. Journal of the American Statistical Association. 1960;55: 698–707.
20. Frees WE, Valdez AE. Understandin relationships using copulas. 32nd. Actuarial Research Conference, 6-8 August at University of Calgary, Albert Canada;1997.
21. Mantel N, Ciminera JL. Use of log rank series in the analysis of litter-matched data on time to tumour appearance. Cancer Research. 1979;39: 4308–4315.
22. Ying Z, Wei LJ. The kaplan-meier estimate for dependent failure time observations. Journal of Multivariate Analysis. 1994;50:17–29.

Peer-review history:
The peer review history for this paper can be accessed here:
http://www.sdiarticle4.com/review-history/66579