Periodicals of Engineering and Natural Sciences ISSN 2303-4521
Vol.8, No.1, March 2020, pp.205-220

A proposed method to estimate the value of dependency for the copula function

Inas Abdulameer Abbood[1], Ismael Hadi Chaloob[2]
[1] Al-Nahrain University, College of Engineering, Baghdad, Iraq
[2] Technical College of Management/Baghdad, Iraq

ABSTRACT

Most of life applications consist of several variables, and for such variables, there should be a relation, which in most cases are complicated, and affect each other. We utilize the association function to find a mathematical relationship among those variables. Association function is considered one as non-parametric functions that used to find the common distributions between the variables in order to model the survival function of two variables. In our research, we use three different relations of Gumbel, Clayton, and Archimedean for Rayleigh distribution of one parameter. Using the association function can possibly estimate the harmony and alignment which achieved between among the variables of the study and from using three methods for estimating the harmony values through the correlation between association function and Kendall correlation function for each particular association. Various values for Rayleigh distribution have been tested with different samples and different experiments in order to find the best harmony value (θ) which represents the rate of optimum association, and to find the binomial distribution. After performing the comparisons by using (MSE), the results show that Clayton Association is the best one for all samples. This association has been selected to be implemented for real data which are selected from Babylon Tires Factory in Najaf governorate for a sample of (50) observations of locally manufactured tires for a variable (L), with operating time measured in minutes (S) for a number of times of speed acceleration where the reliability values have been calculated for each case.

Keywords: Copula function, Rayleigh distribution, Simulation algorithm, Joint distribution, Survival function, Kendall's tau

Corresponding Author:

Inas Abdulameer Abbood
College of Engineering
Al-Nahrain University
E-mail: aenasabdalameer@gmail.com

1. Introduction

Studying copulas and its applications as well as the role that they play in terms of probability is considered as recent phenomenon. Until recently it was difficult to identify the word (copulas), so that it is a tool that used to find the Bivariate Distribution with range [0,1]. This function opens the doors toward a big growing in many fields such as health, management, finance, and industries. Currently the copula function witnesses a fast growing in terms of theoretical and empirical aspects. And now it can be say that copula is a good option to deal with joint distribution easily and accurately if compared with traditional methods.

Naturally, every phenomenon has several variables when be applied, and it does not make sense to consider those variables as independent, so studying these variables requires the formulation a mathematical formula that join these variables. So that joint distribution is considered as a method to join the variables but it accompanied with noticeable time and effort, especially with a huge data. Therefore, the need to find a formula that join such variable raised in order to be used with binomial distributions easily, simply, and more accurately if compared with traditional methods.

1.1. Research problem

It is obviously to find a mathematical formula for joining various variables and find joint distributions, especially if the variables multiply. And to overcome such difficulty, copula function has been utilized to find the desired formula and get the joint distribution.

1.2. Research objectives

This research aims to find relationships with function form that express the context of any two variables which has a high dependency between them, and how to build a common distribution, in addition to formulate a joint distribution for those variables using mathematical functions through single distribution, where a three functions have been used for this purpose in addition to simulation to select the best relation function throughout dependency value estimation ($\hat{\theta}$) which has the least (MSE). After selection the best relation, the function is applied on real data for two variables after selecting single Rayleigh distribution and estimating the survival time and reliable function

1.3. Related work

The first indication for the copula function belongs to Sklar in the context of mathematical statistics, in the following sections, we will describe some related works that are related to our research.

In 2014, Wiboonpongse, Liu, Sriboonchitta and Denoux used nine copulas to estimate the dependency using the simulation, their results stated that the best function was Clayton Copula by using AIC and BIC standards. The study has been implemented on data set taken from agriculture regions for coffee in the north of Thailand [1].

In 2015, Leili has presented a research to study the copula to find a single and multiple relationships among the variables, the researcher used investigation of climate changes and their effect over the drought in producing the biggest five crops in Australia (Wheat, beans, canola oil, lupine and barley) during the period (1980-2012). A model has been formulated using copulas to study the effect for such climate changes, the designed model was intended to mathematically describe the relationship between rain density and crops production in addition to other related factors [2].

In 2016, Fischer has used the copula to count the individual risks and their effect over the total wallet risks to calculate the common binomial distribution among the variables of the study, and to illustrate the role of single variable effect over the total variables through estimating the risk value which represent the single distribution function [3].

In 2016, Gómez, Ausín, and Domínguez have presented a binomial variable model from copulas to find the common distribution for financial stock returning for the development bank of Singapore, in addition to stock circulating in many Asian countries during the period (2010-2014). The researchers relied on copula as a method to identify the causes of benefits reduction for financial returning upon the commercial stocks [4].

In 2017, Kojadinovic has presented a model to distribute a bivariate continuous random variable using copula functions depending on simulation with (1000) frequencies. Assuming four values for Kendall correlation, where the best value was selected through the simulation process to be applied for a set of real data related to several insurance companies. The results proved that using copula is one of the best methods to measure the dependency [5].

A study presented by Dalessandro and Peters, suggested using the copula function with another type of variables which is the discrete variables, the study involves investigation of multiple Markov series, in addition to suggest new formulas to measure the dependency for more than one operation using marginal functions [6].

A model for engineering systems reliability with multiple variables has presented by the researchers of [7]. The authors used the copula function to measure the failure at certain level of shocking loads for system integrity, the study used Micro engine analysis approved by Sandia National Laboratories (SNL), in addition to use number of copulas models. The method proved that Gumball copula present the best performance and most accurate in terms of describing the dependency of margin distributions [7].

1.4. Joint distribution

It is the probability function that gather a number of random variables at the same time, let assume $(y_{1......,}y_m)$ are random variables, then the probability mathematical model describes the variables behavior altogether, the common probability distribution for multiple variables is known as (bivariate distribution), which can be identified using the following formula [8]:

$$F\left(y_{1,........,}y_m\right) = pr[Y_i \leq y_i\,; i = 1, \dots\dots, m] \quad (1)$$

1.5. Marginal distribution [8, 9]

The marginal distribution based on the bivariate distribution, so if we assume that $(y_{1,}y_{2,......,}y_i)$ are random variables with Joint probability distribution function for any given single or two variables, at that time the function $p(y_i)$ or $f(y_i)$ is called marginal function for y_i.

If y_2 , y_1 are two random variables, then the function $p(y_1, y_2)$ or$f(y_1, y_2)$ is called bivariate distribution for the variables (y_1, y_2), and this means:

$$p(Y_1 \leq y_1, Y_2 \leq y_2) = p(Y_1 \leq y_1)p(Y_2 \leq y_2)$$

$$f(y_1, y_2) = f_1(y_1)f_2(y_2) \quad (2)$$

The marginal function has a set of conditions which is related to the Joint distributions, which can be depicted as follows:

1) $f(y)$ is a single-valued function, i.e. for each $x \in \Omega$, there is a single value for $f(y)$.
2) $f(y)$ is always positive value, and for all values, so that the function graph normally located on the X axis.
3) The summation or the integration of the function is equal to one.

1.6. Cumulative distribution function [8]

It is known as the value of the accumulative probability until a given value from the random variable (y), which is identified in the sample space(Ω), it also denoted as $F(y)$, and it is known as the following relationship:

$$\mathrm{F(y) = Pr(Y \leq y)} = \int_{-\infty}^{\mathrm{y}} \mathrm{f(u)du} \; (3)$$

The characteristics of the distribution function for the continuous variables are:

1) $lim_{y\to-\infty} F(y) = F(-\infty) = \int_{-\infty}^{-\infty} f(w)dw = 0$

$$\lim_{y\to\infty} F(y) = F(+\infty) = \int_{-\infty}^{\infty} f(w)dw = Pr,(\Omega) = 1$$

2) The probability of the variable occurrence is specific period can be expressed as:

$$P(a < Y < b) = \int_{a}^{b} \mathrm{f(y)dy} = \int_{-\infty}^{b} \mathrm{f(y)dy} - \int_{-\infty}^{a} \mathrm{f(y)dy}$$

$$= F(b) - F(a)$$

For instance, f(y) is all about a ration of (y) with f(y) which represent the function derivative function

$$\frac{dF(y)}{dy} = f(y) \to dF(y) = f(y)dy$$

1.7. Copula

The copula is known as a multivariate variables distribution function with the domain $[0,1]^n$, it has a number of characteristics which is similar to the bivariate distributions [10, 11]:

1) For each *vu in i:*

$$C(u,0) = 0 = C(0,v)$$

$C(u,1)$ $=u$and $C(1,v)$=v

2) For each u_1, u_2, v_1, v_2 in I, where $u_1 \leq u_2$ and $v_1 \leq v_2$

$$C(u_2,v_2) - C(u_2,v_1) - (u_1,v_2) + C(u_1,v_1) \geq 0$$

1.8. Rayleigh distribution [12, 13]

It is one of the continuous probability distributions, it has a wide range of applications in the context of survival estimation, also in health aspects, especially when dealing with cancer diseases, in addition to communications, and in determining the electronic signal path, in addition to physic sciences such as: measuring wind speed, wave's height, and light ray in engineering. Represented by probability density function (pdf) as follows:

$$f(x;\sigma) = \frac{x}{\sigma^2} e^{-\frac{x^2}{2\sigma^2}} \quad ,x \geq 0 \quad ....(6)$$

Where σ represents the measurement parameter, and the CdF:

$$F(x;\sigma) = 1 - e^{-\frac{x^2}{2\sigma^2}} \quad for\ x \in [0,\infty] ....(7)$$

1.9. Survival function [14,15]

Survival function can be represented by the time period of events until a specific one or more factors occor which cause the death for the biological creatures, mechanic systems failure. This function is called the reliability analysis in engineering, and in economic it is called the perion analysis.

Assuming (T) is denoted for the life time for a particular system, so that the survival of that system in a particular moment (t), is identified by the following formula:

$$R(t) = P(T > t) (8)$$

The survival at a specific moment (t) is called R(t) or survival function, this function can be described according to the cumulative probability distribution function for the random variable (T) as follows:

$$R(t) = 1 - P(T \leq t) = 1 - F(t)$$
$$R(t_1, t_2) = 1 - F(t_1, t_2) (9)$$

Where $F(t_1, t_2)$ is cumulative probability distribution function

1.10. Kendall's tau [16, 17, 18, 19]

It is a non-parametric scale which is used to measure the relation between two pairs of ordinal data, it takes values in the range [1, -1], it also described as the following formula:

$\tau (X; Y) = P[(x_2 - x_1)(y_2 - y_1) > 0] - P[(x_2 - x_1)(y_2 - y_1) < 0]$

which also known as the following form:

$$\tau = \frac{C - D}{C + D} (10)$$

where C= Total Concordant pairs

D= Total Discordant pairs

Let (u, v) is a vector of continuous random variables for the copula C, so it can be say that Kendall's tau as a function to the copula $C(u, v)$ according to the following formula: [13,18]

$$\tau = 4 \int_0^1 \int_0^1 C(u, v) c(u, v) du dv - 1 \ (11)$$

Since that $C(u, v)$ is the CdF for the copula, $c(u, v)$ is the PdF of the copula, and let (x, y) are random variable of the copula C, φ(t) is the derivativefunction that generate the copula, and then Kendall's tau for the copulas is represented by the following formula [15]:

$$\tau = 1 + 4 \int_0^1 \frac{\varphi(t)}{\acute{\varphi}(t)} dt (12)$$

1.11. Types of copulas

1.11.1. Gumbel copula

It is one of the non-parametric function of range [0,1] ^{n}, where the cumulative distributed function has [20,21,22]:

$$C(u, v) = e^{-[(-lnu)^{\theta} + (-lnv)^{\theta}]^{\frac{1}{\theta}}} 1 \leq \Theta < \infty (13)$$

And the p.d.f is

$$c(u, v) = \frac{(lnu\ lnv\ )^{\theta-1}[(-lnu)^{\theta} + (-lnv)^{\theta}]^{\frac{1}{\theta}-2}[(\theta - 1) + ((-lnu)^{\theta} + (-lnv)^{\theta})^{\frac{1}{\theta}}]}{uv\ e^{[(-lnu)^{\theta} + (-lnv)^{\theta}]^{\frac{1}{\theta}}}} \qquad (14)$$

To get the Kendall's correlation for Gumbel copulas, we use the following formula [17,9]:

where $\varphi(t) = (-lnt)^{\theta}$ is the generator function of copula [19], and $\acute{\varphi}(t)$ represents the derivative function, where τcan be calculated by the following formulas:

$$\acute{\varphi}(t)=\frac{-\theta(-lnt)^{\theta-1}}{t} \quad (15)$$

$$=\frac{tlnt}{\theta} \quad (16)$$

$$\tau = 1 + 4\int_0^1 \frac{tlnt}{\theta}dt$$

$$\tau = 1 - \frac{1}{\theta} \quad (17)$$

To get the value of $\hat{\theta}$ which represents the value of the copula, we substitute by the formula (10)

$$\frac{C-D}{C+D} = 1 - \frac{1}{\theta}$$

$$\hat{\theta} = \frac{(C+D)}{2D} \quad (18)$$

After getting the value of $\hat{\theta}$, we estimate the survival function using formula (9) as follows:

$$R(t_1, t_2) = 1 - F(t_1, t_2)$$

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

where:

$$v = 1 - e^{-\frac{y^2}{2\sigma 2^2}} \quad (19)$$

$$u = 1 - e^{-\frac{x^2}{2\sigma 1^2}} \quad (20)$$

$$R(t_1,t_2) = 1 - \left(e^{-\left[\left(-ln\left(1-e^{-\frac{t1^2}{2\sigma 1^2}}\right)\right)^{\theta} + \left(-ln\left(1-e^{-\frac{t2^2}{2\sigma 2^2}}\right)\right)^{\theta}\right]^{\frac{1}{\theta}}} \right) \quad (21)$$

1.11.2. Clayton copula

It is a non-parametric function with domain $[0,1]^n$, since that the CdF with the following formula [16,22,23,24]:

$$C(u,v) = \left(u^{-\theta} + v^{-\theta} - 1\right)^{\frac{-1}{\theta}} \quad 0 < \theta < \infty (22)$$

And PDF for that copula is represented by the following formula:

$$c(u,v) = (vu)^{-(\theta+1)}(1+\theta)\left(u^{-\theta} + v^{-\theta} - 1\right)^{-\frac{1}{\theta}-2} \quad (23)$$

Where: $\varphi(t)$ is the producer function $\varphi(t) = \frac{1}{\theta}\left(t^{-\theta} - 1\right)$

To get the Kendall correlation for Clayton copulas and according to the formula [12]:

$$\acute{\varphi}(t) = -t^{-\theta-1} (24)$$

$\acute{\varphi}(t)$ is the derivative function of the copula:

$$\tau = 1 + 4\int_0^1 \frac{\varphi(t)}{\acute{\varphi}(t)}dt \qquad 0 < t < 1$$

$$\tau = \frac{\theta}{\theta + 2} (25)$$

To get the value of $\hat{\theta}$, which represents the value of the copula, then we substitute by the equation (10).

$$\frac{C-D}{C+D} = \frac{\theta}{\theta+2}$$

$$\hat{\theta} = \frac{C-D}{D} (26)$$

After finding the value of $\hat{\theta}$, then we estimate the survival function from equation (9):

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

We substitute for every (u) with the equation (19), and for each (v) with equation (20)

$$R(t_1, \mathrm{t}_2) = 1 - \left[\left(1 - e^{-\frac{t1^2}{2\sigma1^2}}\right)^{-\theta} + \left(1 - e^{-\frac{t2^2}{2\sigma2^2}}\right)^{-\theta} - 1\right]^{\frac{-1}{\theta}} (27)$$

$$= 1 + 4\int_0^1 \frac{\left(1 - t^{\frac{1}{\theta}}\right)^{\theta}}{\left(1 - t^{\frac{1}{\theta}}\right)^{\theta-1} - t^{\frac{1}{\theta}-1}} dt$$

$$\tau = \frac{(2\theta - 3)}{(2\theta - 1)} (28)$$

To get the estimated value of θ, which represents the copula, we substitute with equation (10).

$$\frac{(C-D)}{(C+D)} = \frac{(2\theta-3)}{(2\theta-1)}$$

$$\hat{\theta} = \frac{(C+2D)}{2D} (29)$$

After finding the value of $\hat{\theta}$, we estimate the survival function from equation (9), as follows:

$$R(u,v) = 1 - \left[\left\{\max\left(1 - \left[\left(1 - \left(1 - e^{-\frac{x^2}{2\sigma^2}}\right)^{\frac{1}{\theta}}\right)^{\theta} + \left(1 - \left(1e^{-\frac{y^2}{2\sigma^2}}\right)^{\frac{1}{\theta}}\right)^{\theta}\right]^{\frac{1}{\theta}}, o\right)\right\}^{\theta}\right]^{-\theta} \quad (30)$$

$$C(u,v) = \max\left[1 - \left((1-u)^{\theta} + (1-v)^{\theta}\right)^{\frac{1}{\theta}}\right] \theta \in [1, \infty) \ (31)$$

1.11.3. Archimedean 3 copula [25]

The CdF for Archimedean 3 copula is described as below:

$$C(u,v) = \max\left[1 - \left((1-u)^{\theta} + (1-v)^{\theta}\right)^{\frac{1}{\theta}}\right] \theta \in [1, \infty) \ (32)$$

And the generator function for the copula is $\varphi(t) = (1-t)^{\theta}$, and to get the Kendall correlation we follow the formula (12) as illustrated below:

$\hat{\varphi}(t) = -\theta(1-t)^{\theta-1}(33)$

$\tau = 1 + 4\int_0^1 \frac{(1-t)^{\theta}}{-\theta(1-t)^{\theta-1}}dt$

$\tau = \frac{\theta - 2}{\theta} \ldots\ldots\ldots (34)$

To find the estimated value of θ which represents the copula, then we substitute by formula (10) as below:

$$\frac{C-D}{C+D} = \frac{\theta+2}{\theta}$$

$$\hat{\theta} = \frac{(D\ +c)}{D} \ldots..(35)$$

After finding θ, then the survival function is estimated as the following processes:

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

Each $\boldsymbol{u}$ is substitiuted by formula (19), and each $\boldsymbol{v}$ is substituted by formula (20)

$$\boldsymbol{R}(\boldsymbol{t_1},\mathbf{t_2}) = \mathbf{1} - \mathbf{max}\left[\mathbf{1} - \left(\left(\mathbf{1} - \left(\mathbf{1} - \boldsymbol{e}^{-\frac{t_1{}^2}{2^{\sigma}1^2}}\right)\right)^{\boldsymbol{\theta}} + \left(\mathbf{1} - \left(\mathbf{1} - \boldsymbol{e}^{-\frac{t_2{}^2}{2\sigma 2^2}}\right)\right)^{\boldsymbol{\theta}}\right)^{\frac{1}{\theta}}\right] \ldots (\mathbf{36})$$

$$\boldsymbol{R}(\boldsymbol{t_1},\boldsymbol{t_2}) = \mathbf{1} - \left(\left[\left(\mathbf{1} - \boldsymbol{e}^{-\frac{t1^2}{2\sigma 1^2}}\right)\left(\mathbf{1} - \boldsymbol{e}^{-\frac{t2^2}{2\sigma 2^2}}\right)\right]\left[\mathbf{1} + \boldsymbol{\theta}\left(\mathbf{1} - \left(\mathbf{1} - \boldsymbol{e}^{-\frac{t1^2}{2\sigma 1^2}}\right)\right)\left(\mathbf{1} - \left(\mathbf{1} - \boldsymbol{e}^{-\frac{t2^2}{2\sigma 2^2}}\right)\right)\right]\right) \ldots (\mathbf{37})$$

2. Simulation

The vast development in the context of information technology is considered as new trend that courage many researchers to utilize the simulations in terms of practical representation for a particular researches that aim to study the parameters behavior or statistical distributions, so that the simulation is the process of imitating the real, physical, or biological performance for a particular system behavior or to study a model in order to build a real system. The simulation aims to study the potential results and to get an information or expectations in the future in order to save time, effort, and money. Also this process allows forecasting about production behavior in certain circumstances where the simulation can imitate different operation that the real system can do, hence this can help to test different theories and hypothesis.

2.1. Building a simulation program stages

As mentioned before, the simulation is all about imitating for the reality in order to estimate a new parameter, so it can be used to estimate the copula parameters, survival function, and the reliability for Rayleigh distribution as the following:

a. The stage of identifying the default values

b. The stage of generating the random data to perform the simulation for two variables (x, y) by using the formula (9) as illustrated below:

$$\boldsymbol{F}(\mathbf{x};\boldsymbol{\sigma}) = \mathbf{1} - \boldsymbol{e}^{-\frac{x^2}{2\sigma^2}}$$

$$\boldsymbol{u} = \mathbf{1} - \boldsymbol{e}^{-\frac{x^2}{2\sigma^2}}$$

$$x = \sigma\sqrt{-2\ln(1-u)} \ldots\ldots\ldots\ldots \sigma > 0 \ldots\ldots\ldots (38)$$

And by using same way with variable (y):

$$y = \sigma\sqrt{-2\ln(1-v)} \ldots\ldots\ldots \sigma > 0 \ldots\ldots\ldots (39)$$

2.2. Mean square error (MSE)

Is a scale used to compare the preference, which is widely used in accurately prediction, it represents the average of square differences between the real and the expected parameters, MSE can be described as the following formulas[26]:

$$MSE = E(\theta - \widehat{\theta}_i)^2 \ldots\ldots\ldots (40)$$

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(\widehat{\theta} - \theta)^2 \ldots\ldots (41)$$

Where: n is the frequencies for each experiment, $\widehat{\theta}$ is the estimated value, θ is the real value

3. Results and discussion

3.1. Simulation results

The results of the simulation will be displayed and analyzed to estimate the dependency value between the variables ($\widehat{\theta}$) according to the method mentioned in the literatures of Rayleigh distribution, where the experiment was repeated (1000) times. The following Tables illustrate the simulation results.

- Table (1) describes the simulation results for the first experiment with sample size (n=25) and dependency value ($\theta = 2$). As listed in the Table, we note that the Clayton copula has the best MSE, and the more values was at the Archimedean 3 Copula.
- In Table (2) and at the sample size (n=50) under the selected generated distribution with (1000) frequencies, according to the estimation method for the copulas mentioned before, and at dependency value ($\theta = 2$), we note that the MSE for the three copulas was in Clayton copula, however, the biggest and the more values was in Archimedean 3 Copula. The best estimation values occurs for two copulas for the same sample size under the same conditions for the parameters $\sigma_1 = 0.8, \sigma_2 = 0.3$.
- In Table (3), with sample size (n=75) under generated Rayleigh distribution, and with (1000) frequencies, we note that the best estimation value comes with Clayton copula, the more values occurs with Archimedean 3 Copula.
- From Table (4), and at sample size (n=100), which is the biggest assumed sample size, and under the same conditions, the best estimations value for the three copulas was in Clayton copula, however, the biggest and the more values was in Archimedean 3 Copula, and the least MSE for the parameters ($\sigma_1 = 0.8, \sigma_2 = 0.8$) was in Archimedean3 Copula.

Table 1. The MSE in Experiment (1) when (θ=2)

n	σ_1	σ_2	M_1 Gumbel	M_2 Clayton	M_3 Archimedean3
	0.1	0.1	1.6887931	0.0108844	3.7357669
	0.1	0.3	1.6935213	0.0096525	3.7496436
	0.1	0.5	1.6904059	0.0097252	3.7403156
	0.1	0.8	1.6956111	0.0105718	3.7561430
	0.3	0.1	1.5765987	0.0279665	3.4034542
	0.3	0.3	1.5713984	0.0285979	3.3880114
	0.3	0.5	1.5714795	0.0284710	3.3882230

n	σ_1	σ_2	M_1 Gumbel	M_2 Clayton	M_3 Archimedean3
25	0.3	0.8	1.5832186	0.0260310	3.4228301
	0.5	0.1	1.4795479	0.0561216	3.1193407
	0.5	0.3	1.4796463	0.0561670	3.1196474
	0.5	0.5	1.4777799	0.0568981	3.1142309
	0.5	0.8	1.4787359	0.0564166	3.1169786
	0.8	0.1	1.4401575	0.0715942	3.0050376
	0.8	0.3	1.4420972	0.0706840	3.0106292
	0.8	0.5	1.4406382	0.0712987	3.0064060
	0.8	0.8	1.4422575	0.0706368	3.0110983

Table 2. The MSE in Experiment (1) and medium sample size, when (θ=2)

n	σ_1	σ_2	M_1 Gumbel	M_2 Clayton	M_3 Archimedean3
	0.1	0.1	1.6896593	0.0066537	3.7373079
	0.1	0.3	1.6918340	0.0063943	3.7437673
	0.1	0.5	1.6869616	0.0067018	3.7292269
	0.1	0.8	1.6912704	0.0063170	3.7420573
	0.3	0.1	1.5694990	0.0269515	3.3819016
	0.3	0.3	1.5753886	0.0254303	3.3991900
	0.3	0.5	1.5713208	0.0265707	3.3872717
	0.3	0.8	1.5698869	0.0268086	3.3830295
	0.5	0.1	1.4753601	0.0567784	3.1069416
	0.5	0.3	1.4756044	0.0566507	3.1076426
50	0.5	0.5	1.4750068	0.0568696	3.1059046
	0.5	0.8	1.4761869	0.0564498	3.1093399
	0.8	0.1	1.4393951	0.0715115	3.0027298
	0.8	0.3	1.4390969	0.0716643	3.0018734
	0.8	0.5	1.4395060	0.0714639	3.0030508
	0.8	0.8	1.4391093	0.0716479	3.0019065

Table 3. The MSE in Experiment (1) and medium sample size when (θ=2)

n	σ_1	σ_2	M_1 Gumbel	M_2 Clayton	M_3 Archimedean3
	0.1	0.1	1.6892092	0.0056740	3.7357126
	0.1	0.3	1.6873131	0.0059925	3.7301041
	0.1	0.5	1.6899856	0.0057864	3.7380702
	0.1	0.8	1.6908526	0.0057780	3.7406690
	0.3	0.1	1.5722961	0.0258700	3.3900224
	0.3	0.3	1.5713997	0.0260449	3.3873770
	0.3	0.5	1.5720682	0.0258706	3.3893389
	0.3	0.8	1.5725311	0.0258007	3.3907101
	0.5	0.1	1.4751555	0.0566747	3.1063020
75	0.5	0.3	1.4750571	0.0567187	3.1060175
	0.5	0.5	1.4753926	0.0565683	3.1069865
	0.5	0.8	1.4756826	0.0564812	3.1078347

n	σ1	σ2	M_1 Gumbel	M_2 Clayton	M_3 Archimedean3
	0.8	0.1	1.4413868	0.0705762	3.0084710
	0.8	0.3	1.4409913	0.0707500	3.0073280
	0.8	0.5	1.4410631	0.0707161	3.0075349
	0.8	0.8	1.4410673	0.0707274	3.0075505

Table 4. The MSE in Experiment (1) and big sample size when (*θ=2*)

n	σ1	σ2	M_1 Gumbel	M_2 Clayton	M_3 Archimedean3
100	0.1	0.1	1.6876824	0.0054984	3.7310886
	0.1	0.3	1.6892263	0.0053242	3.7356767
	0.1	0.5	1.6891373	0.0053218	3.7354089
	0.1	0.8	1.6893823	0.0053068	3.7361402
	0.3	0.1	1.5700635	0.0262247	3.3834134
	0.3	0.3	1.5703706	0.0261092	3.3843059
	0.3	0.5	1.5717173	0.0257861	3.3882652
	0.3	0.8	1.5717873	0.0257801	3.3884735
	0.5	0.1	1.4749365	0.0566969	3.1056504
	0.5	0.3	1.4760307	0.0562775	3.1088281
	0.5	0.5	1.4752707	0.0565765	3.1066228
	0.5	0.8	1.4750934	0.0566404	3.1061070
	0.8	0.1	1.4412804	0.0704149	3.0093114
	0.8	0.3	1.4403397	0.0701229	3.0112166
	0.8	0.5	1.4424902	0.0700607	3.0116524
	0.8	0.8	1.4410522	0.0703281	3.0058954

- Table (5) describes the second experiment for the simulation with sample size (n=25) which is the smallest size, and with dependency ($\theta = 4$), and (1000) times of frequencies. The best estimation under this sample for the three copulas comes with Clayton copula, the more values occurs with Archimedean 3 Copula.
- In Table (6), the sample size (n=50) and (1000) frequencies and dependency $\theta = 4$, The best estimation under this sample for the three copulas comes with Clayton copula, the more values occur with Archimedean 3 Copula, also the best estimation values occurs for the parameters ($\sigma_1 = 0.8, \sigma_2 = 0.1$).
- In Table (7) and with a sample size (n=75) under the same conditions, the best estimation under this sample for the three copulas comes with Clayton copula, the more values occur with Archimedean 3 Copula.
- In Table (8), the sample size (n=100) which is the largest size, and (1000) frequencies and dependency $\theta = 4$, The best estimation under this sample for the three copulas comes with Clayton copula, the more values occur with Archimedean 3 Copula, also the best estimation values occurs for the parameters ($\sigma_1 = 0.8, \sigma_2 = 0.3$)

Table 5. The MSE in Experiment (2) and small sample size, when (*θ=4*)

n	σ1	σ2	M_1 Gumbel	M_2 Clayton	M_3 Archimedean3
	0.1	0.1	2.0779138	0.0126292	4.3405050
	0.1	0.3	2.0913203	0.0159757	4.3797653

n	σ_1	σ_2	M_1 Gumbel	M_2 Clayton	M_3 Archimedean3
	0.1	0.5	2.0784563	0.0127938	4.3421019
	0.1	0.8	2.0875031	0.0152355	4.3686476
	0.3	0.1	1.9519889	0.0029116	3.9779429
	0.3	0.3	1.9553880	0.0035269	3.9878305
	0.3	0.5	1.9640269	0.0045042	4.0127922
25	0.3	0.8	1.9548098	0.0035152	3.9861751
	0.5	0.1	1.8363407	0.0096151	3.6494349
	0.5	0.3	1.8409732	0.0092338	3.6625617
	0.5	0.5	1.8398638	0.0091185	3.6593589
	0.5	0.8	1.8411824	0.0089730	3.6630849
	0.8	0.1	1.7874886	0.0164032	3.5117968
	0.8	0.3	1.7906038	0.0158452	3.5205380
	0.8	0.5	1.7809057	0.0159945	3.5256390
	0.8	0.8	1.7906038	0.0158452	3.5245790

Table 6. The MSE in Experiment (2) and medium sample size, when (*θ=4*)

n	σ_1	σ_2	M_1 Gumbel	M_2 Clayton	M_3 Archimedean3
	0.1	0.1	2.0813282	0.0091655	4.3492706
	0.1	0.3	2.0805404	0.0093843	4.3470824
	0.1	0.5	2.0824972	0.0098280	4.3528001
	0.1	0.8	2.0788005	0.0092158	4.3420631
	0.3	0.1	1.9512517	0.0012158	3.9753522
	0.3	0.3	1.9524925	0.0013528	3.9789364
	0.3	0.5	1.9519826	0.0012535	3.9774512
	0.3	0.8	1.9527552	0.0012506	3.9796578
	0.5	0.1	1.8374121	0.0084694	3.6521685
	0.5	0.3	1.8356576	0.0087383	3.6472328
	0.5	0.5	1.8344237	0.0087503	3.6437106
50	0.5	0.8	1.8361371	0.0085535	3.6485499
	0.8	0.1	1.7904971	0.0154997	3.5201344
	0.8	0.3	1.7917282	0.0152949	3.5235935
	0.8	0.5	1.7906007	0.0154568	3.5204182
	0.8	0.8	1.7927568	0.0154979	3.5255785

Table 7. The MSE in Experiment (2) and medium sample size, when (*θ=4*)

n	σ_1	σ_2	M_1 Gumbel	M_2 Clayton	M_3 Archimedean1
	0.1	0.1	2.0789069	0.0082925	0.8871806
	0.1	0.3	2.0786743	0.0083168	0.8870332
	0.1	0.5	2.0798126	0.0083413	0.8877672
	0.1	0.8	2.0794963	0.0084679	0.8875751
	0.3	0.1	1.9534613	0.0007504	0.8058636
	0.3	0.3	1.9509087	0.0007359	0.8042213

n	σ_1	σ_2	M_1 Gumbel	M_2 Clayton	M_3 Archimedean1
75	0.3	0.5	1.9504609	0.0006441	0.8039252
	0.3	0.8	1.9528382	0.0007440	0.8054624
	0.5	0.1	1.8364371	0.0083519	0.7313124
	0.5	0.3	1.8360434	0.0083987	0.7310635
	0.5	0.5	1.8354644	0.0084461	0.7306955
	0.5	0.8	1.8355263	0.0084878	0.7307391
	0.8	0.1	1.7933285	0.0148954	0.7041840
	0.8	0.3	1.7927072	0.0150218	0.7037959
	0.8	0.5	1.7928223	0.0149949	0.7038675
	0.8	0.8	1.7932633	0.0148765	0.7048665

Table 8. The MSE in Experiment (2) and big sample size, when (*θ=4*)

n	σ_1	σ_2	M_1 Gumbel	M_2 Clayton	M_3 Archimedean3
	0.1	0.1	2.0786148	0.0078962	4.3411555
	0.1	0.3	2.0792773	0.0078478	4.3430344
	0.1	0.5	2.0782818	0.0078140	4.3401806
	0.1	0.8	2.0805522	0.0080461	4.3467339
	0.3	0.1	1.9514800	0.0005283	3.9758082
	0.3	0.3	1.9530498	0.0005688	3.9803048
	0.3	0.5	1.9525376	0.0005731	3.9788427
	0.3	0.8	1.9526125	0.0005734	3.9790566
	0.5	0.1	1.8361646	0.0083580	3.6485724
	0.5	0.3	1.8356745	0.0083796	3.6471786
	0.5	0.5	1.8356057	0.0083510	3.6469736
100	0.5	0.8	1.8347113	0.0084934	3.6444591
	0.8	0.1	1.7944493	0.0146594	3.5311863
	0.8	0.3	1.7940357	0.0147599	3.5300333
	0.8	0.5	1.7944649	0.0146703	3.5312340
	0.8	0.8	1.7940379	0.0147300	3.5303534

3.2. Empirical results

After reaching the best copula using the simulation, we use the best parameter to get the bivariate probability distribution, the next sections depict the empirical side of for measuring the reliability function in the industrial sector for Babylon Tiers Factory. To apply the best copula function to find the survival function for the common binomial distribution for Rayleigh distribution, then the hypothesis should be tested which follow the following steps:

1- Identify the null hypothesis and the alternative one.
   - H_0: It assumes that the observations group follow the selected distribution.
   - H_1: It assumes that the observations group do not follow the selected distribution.
2- Identify the level of significance ($\alpha = 0.05$), which is the possibility of rejecting H_0, when it is correct.
3- Perform good matching for the data using Chi-Square test of independency.
4- Identify select the value $\theta = 4$ for the copula with matches the least MSE in Clayton copula simulation, and assign estimation values for the one-way distribution$\sigma_1 = 0.3, \sigma_2 = 0.1$.
5- Calculate the survival function R (X, y) using formula (9).

We apply our method in Babylon Tiers Factory, at Najaf Governorate. The factory has been designed to product cars tiers with production capacity of two million tires and more yearly, the purpose of this factory was to cover the country needs with the ability of exporting the products to the nearby countries, the factory produces almost all standard sizes. However, due to the bad circumstances that Iraq faced lately prevented providing some tools and machines from the suppliers, the thing that forced the factory to run with what is available and with the least production capacity. The data have been recorded from the past registered records for the years (2012-2013), where 50 observations were selected for two types of local tires, the tire was exposed to different speeds, so we change the speed each 10 minutes for hours, with pressure stability with 85%, where we the first variable was the starting time denoted by (L), the second variable is the speed increment (s). From the formula (9), we come out with (R (L, S)), as described in the Table (9).

Table 9. Reliability values calculation

Seq.	Strat time in minutes (L)	Speed increment times (S)	R(L,S)	Seq	Strat time in minutes (L)	Speed increment times (S)	R(L,S)
1	32	4	0.8252	26	70	7	0.6848
2	52	5	0.7724	27	51	6	0.7252
3	59	6	0.7243	28	61	7	0.6844
4	52	6	0.7254	29	68	7	0.6841
5	45	5	0.7728	30	89	9	0.6171
6	35	4	0.8252	31	55	6	0.7254
7	45	5	0.7728	32	62	7	0.6845
8	45	5	0.7728	33	58	6	0.7254
9	42	6	0.7254	34	56	6	0.7253
10	53	6	0.7250	35	60	6	0.7254
11	82	9	0.6171	36	61	7	0.6844
12	51	6	0.7252	37	50	5	0.7715
13	57	6	0.7250	38	41	5	0.7718
14	55	6	0.7254	39	61	7	0.6844
15	42	5	0.7720	40	11	2	0.9532
16	65	6	0.7254	41	59	6	0.7253
17	35	4	0.8252	42	53	6	0.7254
18	60	6	0.7254	43	45	5	0.7728
19	44	5	0.7717	44	55	6	0.7254
20	60	6	0.7254	45	12	2	0.9531
21	101	11	0.5627	46	55	6	0.7254
22	122	13	0.5185	47	44	5	0.7717
23	129	13	0.5183	48	63	7	0.6842
24	80	8	0.6491	49	67	7	0.6846
25	88	8	0.6490	50	65	7	0.6845

The above Table shows that the reliability was (0.5183, 0.5185) for both tiers with sequence (22, 23) versus 13 times of speed increment, whereas the reliability for the tires (40, 45) was (0.9).

4. Conclusions

After performing the research outlines, we come out with a number of conclusions based on the results gained:

1. During the simulation, MSE for Clayton copula presented the best estimations and for the both experiments, as well as for different sample sizes.

2. From using the simulation model, the best three copulas and for all experiments presented the best (MSE), and for different samples and dependencies.
3. For Rayleigh distribution, the first experiment showed a dependency $(\theta = 2)$, and for sample size (n=50,100), and a parameters with ($\sigma_1 = 0.8, \sigma_2 = 0.1$) for Clayton copula.
4. The increase of the (MSE) for the involved sample sizes (n=50,75,100) with increasing the sample size, the Rayleigh distribution have affected the results ($\sigma_1 = 0.8, \sigma_2 = 0.3$) for the Gumbel copula, so that we get the least dependency $(\theta = 2)$.
5. According to the simulation, the largest (MSE) with (4.5218881) for Archimedean 3 copula for the parameters ($\sigma_1 = 0.1, \sigma_2 = 0.5$), with dependency of $\theta = 5$, and sample size (n=25), where the more increase in the sample size under the same conditions, then the parameters reduces, and this matches the statistical theory, where it was (4.5059073) with sample size (n=100).
6. The value of ($\hat{\theta}$) has a big impact in survival function calculation, since it affects in the harmonic ration that occur for the bivariate distribution which extracted from the copula function. From all of this, we can say that the copulas make the dependency relationship between the variables better in terms of explanations and more accurate in terms of mathematical expressions, which make the results more accurate.
7. From all of the results we have got, and generally, it is obvious that the reliability function reduces when the speed increased.
8. The copulas present flexible models that can be implemented in different perspectives.
9. The Archimedean copulas family are considered better scale to measure the dependency and its applications.

5. Recommendations

We present some recommendations that suits the results of our study:

1. We recommend using the copulas since it considered as a starting point for building joint distributions which follow the same domain for other distributions or the distributions with marginal functions.
2. It is possible to use the copulas to study the dependency for more than two variables, which would be of the best traditional methods including multiple variables.
3. Using the copulas to build a joint distribution to get the right behaviour for the variables dependency, so that it is not possible to measure the variables effect unless there is a specific formula to measure the dependency.

References

[1] A. Wiboonpongse, J. Liu, S. Sriboonchitta and T. Denoeux, "Modeling dependence between error components of the stochastic frontier model using copula: application to intercrop coffee production in Northern Thailand" , International Journal of Approximate Reasoning, vol.65, pp. 34-44, 2015.

[2] L. Li , "Integrating an Agro-Meteorological Indicator for Assessing Drought Impacts on Agricultural Production", University of California, Irvine, 2015.

[3] P. Jaworski, F. Durante, W. K. Hardle and T.Rychlik, "Copula theory and its applications" ,2010.

[4] M. Gómez, M. C. Ausín and M. C. Domínguez, "Seasonal copula models for the analysis of glacier discharge at King George Island", Antarctica. Stochastic Environmental Research and Risk Assessment, vol. 31, no.5, pp.1107-1121, 2017.

[5] I. Kojadinovic, "Some copula inference procedures adapted to the presence of ties", Computational Statistics & Data Analysis, vol. 112, pp.24-41, 2017.

[6] A. Dalessandro and G.W. Peters, "Tensor Approximation of Generalized Correlated Diffusions and Functional Copula Operators", Methodology and Computing in Applied Probability, pp.1-35, 2017.

[7] Z. An, Z., D. Sun,.“Reliability modeling for systems subject to multiple dependent competing failure processes with shock loads above a certain level”,Reliability Engineering & System Safety, vol. 157, pp.129-138, 2017

[8] A. H. Hormuz, “Mathematical Statistics”, Ministry of Higher Education / University of Mosul, College of Administration and Economics, 1990.

[9] Marta Monika Wawrzyniak, “Dependence Concepts” , Master Thesis in Risk and Environmental Modelling Written under the direction of Dr Dorota Kurowicka,2006.

[10] M. Fischer , " Tailoring copula-based multivariate generalized hyperbolic secant distributions to financial return data: An empirical investigation”, Germany, Institute of Statistics and Econometrics University of Erlangen-N¨urnberg, LangeGasse , 2003.

[11] H. P. Palaro and L. K. Hotta, “ Using conditional copula to estimate value at risk”,American journal of clinical nutrition, vol.70, no.6,pp. 1001-1008,2004.

[12] E. E. Afify, “Comparison of estimators of parameters for the Rayleigh distribution”, Faculty of Eng. Shibeen El KomMenoufia Univ. articles/0302004. Pdf, 2004.

[13] A. Ahmed, S. P. Ahmad and J.A. Reshi, “Bayesian analysis of Rayleigh distribution”. International Journal of Scientific and Research Publications, vol.3, no.10, pp.1-9, 2013.

[14] L. Taj, A. Sarhan, “Introduction to Random Variables”, King Saud University, College of Science / Department of Statistics and Operations Research, 2011.

[15] J. Fox, “Cox proportional–hazards regression for survival data”, An R and S-PLUS companion to applied regression, 2002. [20]C. R. Bhat, and N. Eluru, “A copula-based approach to accommodate residential self-selection effects in travel behavior modeling”, Transportation Research Part B: Methodological, vol.43, no. 7, pp.749-765, 2009.

[16] A. Charpentier, “Tail distribution and dependence measures ”, International Actuarial Association Proceedings of the 34th ASTIN Conference ,pp.1-25,(2003,March).

[17] Mario R. Melchiori1 , “Credit Derivative Models: Which Archimedean Copula is the right one? ”, Universidad Nacional del Litoral Santa Fe - Argentina Received after revisions September 2003

[18] S. R. Das, andG. Geng, “Simulating correlated default processes using copulas: a criterion-based approach. Working paper”, University of Santa Clara and Gifford Fong Associates, 2003.

[19] Casey Quinn , “Using copulas to measure association between ordinal measures of health and income”, University of York, USA October 22, 2007.

[20] C. R. Bhat, and N. Eluru, “A copula-based approach to accommodate residential self-selection effects in travel behavior modeling”, Transportation Research Part B: Methodological, vol.43, no. 7, pp.749-765, 2009.

[21] Eric Bouye, ValdoDurrleman, AshkanNikeghbali, GaÄelRiboulet& Thierry Roncalli “Copulas for Finance a Reading Guide and Some Applications”, Groupe de Recherche Operationnelle, Bercy-Expo | ImmeubleBercySud ,2000.

[22] R. Radice, G. Marra and M. Wojtyś,“Copula regression spline models for binary outcomes” ,Statistics and Computing, vol. 26, no.5, pp. 981-995,2016.

[23] P. Deb,P.K. Trivedi and D. M. Zimmer, “Dynamic cost-offsets of prescription drug expenditures: Panel data analysis using a copula-based hurdle model”, National Bureau of Economic Research.,2009.

[24] Sel Ly , Kim-Hung Pho, Sal Ly 1 and Wing-Keung Wong, “Determining Distribution for the Product of Random Variables by Using Copulas”, MDPI,7, 23; doi:10.3390/risks7010023,2019.

[25] R. B. Nelsen, “An Introduction to Copula ”, New York: Springer, USA 97219-7899,2006.

[26] Paul, Sanjoy Kumar. “Determination of exponential smoothing constant to minimize mean square error and mean absolute deviation”, *Global journal of research in engineering* 11.3 (2011).