The 3rd Work shop on Copula and Various types of Dependencies

25-26 February 2015

Department of Statistics,
Shahid Bahonar University of Kerman, Iran
Ordered and Spatial Data Center of Excellence,
Ferdowsi University of Mashhad, Iran

copula Theory

- Construction methods
- Modeling using copula
- copula and its application in spatial statistics
- Copula and its application in financial mathematics

Address : Department of statisitics
Faculty of Mathemathics & Computer, Shahid Bahonar University of Kerman
P.O.Box: 76169-14111 ,Kerman.Iran. Tel.Fax: +98 343322 0057
website: http://statistics.fa.uk.ac.ir Email: statisticsdept@mail.uk.ac.ir

In The Name of Allah

Proceedings
(Papers in English)

The 3^{rd} Workshop on Copula and Various Types of Dependencies

Department of Statistics, Faculty of Mathematics and Computer, Shahid Bahonar University of Kerman, Iran

Ordered and Spatial Data Center of Excellence, Ferdowsi University of Mashhad, Iran

25-26 February 2015

Scientific Committee:

Amini, M.	Ferdowsi University of Mashhad
Tata, M. B. (Chairman)	Shahid Bahonar University of Kerman
Dolati, A.	University of Yazd
Rezapour, M.	Shahid Bahonar University of Kerman
Alamatsaz, M. H.	University of Isfahan
Mohamadzadeh, M.	Tarbiat Modares University

Organizing Committee:

Ahmadi, J.	Ferdowsi University of Mashhad
Arabpour, A.	Shahid Baahonar University of Kerman
Jabari Noughabi, H.	Ferdowsi University of Mashhad
Khosravi, M.	Shahid Bahonar University of Kerman
Iranmanesh, R.	Shahid Bahonar University of Kerman
Pourmousa, R. (Chairman)	Shahid Bahonar University of Kerman
Sheikhy, A.	Shahid Bahonar University of Kerman
Taherizaheh, H.	Shahid Bahonar University of Kerman

Topics of the Workshop:

- History of Copula
- Construction of Copula
- Copulas and Dependence Concepts
- Modelling using Copula
- Applications of Copula in Financial Mathematics
- Copula and Spatial Statistics
- Software Concepts of Copula

Preface

On behalf of the organizing and scientific committees, we would like to extend a very warm welcome to all the participants of the 3rd workshop on copula and various types of dependencies. We hope this workshop provides an environment for useful discussion and exchange of scientific ideas.

We wish to express our gratitude to the numerous individuals and organizations that have contributed to the success of this workshop, in which more than 40 colleagues, researchers, and postgraduate students have participated.

Finally, we would like to extend our sincere gratitude to the students of the Department of Statistics at Shahid Bahonar University of Kerman for their unstinting cooperation.

Department of Statistics
Faculty of Mathematics and Computer
Shahid Bahonar University of Kerman

Ordered and Spatial Data
Center of Excellence
Ferdowsi University of Mashhad

Contents

The 3rd workshop on
Copula and various types of dependencies
25-26, February 2015

APPLICATION OF COPULA IN GROUNDWATER QUALITY INTERPOLATION

GANJALIKHANI, M.*[1], ZOUNEMAT-KERMANI, M.[1], RAHNAMA, M. B.[1] AND REZAPOUR, M.[2]

[1] *Department of Water Engineering, Shahid Bahonar University of Kerman, Kerman, Iran;*
Ganjalikhani@agr.uk.ac.ir; zounemat@uk.ac.ir; mbr@uk.ac.ir.

[2] *Department of Statistics, Shahid Bahonar University of Kerman, Kerman, Iran*
mohsenrzp@uk.ac.ir

Abstract. This study presents a new method for interpolation by using copula for groundwater quality zoning. For this purpose the qualitative data of 107 observation wells in Ravar and Kerman plain contains Bicarbonate, Sulphate, Calcium and Total Dissolved Solids (TDS) at winter of 2014 were used. Then, the obtained results were compared to the results obtained from conventional zoning methods to evaluate the performance of copulas. Analysis of the results with respect to the root mean square error showed that copula has a higher ability than common methods in qualitative zoning of groundwater resources.

1. Introduction

In classical statistical analysis, samples have no spatial information in space and consequently, the specific sample is not including any information about the same sample in the found distance. But, geostatistics methods have overcome these challenges. The main weakness of this method has been normal data condition that occurs in natural conditions less than in others. Groundwater often has skews and the assumption of having normal distribution is not respected here. The sensitivity to outliers of this method is another disadvantage. In this study we have tried to present a new method

2010 *Mathematics Subject Classification.* Primary 47A55; Secondary 39B52, 34K20, 39B82.

Key words and phrases. Copula , Groundwater Quality, Geostatistic, Kriging.

* Speaker.

for interpolation by using copula. In this approach, the possibility of taking advantage of all the copula family such as Archimedean family which has a high flexibility for data processing, is available in terms of the groundwater quality. For this purpose, the qualitative data of 107 observation wells in Ravar and Kerman plain contains Bicarbonate, Sulphate, Calcium and Total Dissolved Solids (TDS) at winter of 2014 were used. The obtained results of this method by use of RMSE statistics, were compared with IDW, Kriging, Kriging with Log transformation and Kriging with Box-Cox transformation methods, and the obtained statistics indicated that the accuracy of the presented method is much higher than other methods.

2. Materials and Methods

2.1. **Description of Spatially Variable Structure.** Copula can be used as bivariate distribution function for the points at a fixed distance to describe and analyze the spatial structures. In this regard, the appropriate marginal distribution function is fitted to the data for variables of region Z. Then, the data related to variables which their distance from each other is $h \pm \Delta h$ are separated which in this formula h is desired distance and Δh is amplitude. Usage of amplitude is necessary because of irregular distribution of wells and low probability of finding some paired samples with the same distances. It should be noted that variables conditions of regions must be applied to the data. Just like variogram, as a numeric value is determined for each physical distance, in this method, a copula is fitted for each interval. Consequently, copula has more contains and more information than variogram. So, copula for each paired data with distance of $h \pm \Delta h$ can be considered as the following formula:

$$C_h = C(F_z(Z(x)), F_z(Z(x+h))) \tag{2.1}$$

To describe the spatial structure, the other two conditions must be applied as follows.

1. For $\|h\| \mapsto \infty$ there should be $C_h(u) \mapsto \Pi^2(u)$, and this indicates that, these two variables are completely independent of each other in the large distances.

2. For $\|h\| \mapsto 0$ there should be $C_h(u) \mapsto M^2(u)$, and this condition also implies that, at very short distances the two variables are entirely dependent on each other. So, Nugget effect phenomenon which is created in calculation and drawing of variograms, will be lost. Then, the classification distances is considered for each class of combined copulas, which are obtained by combining of beginning and ending copula of classification distance.

$$C_h(u_1,u_2) := \begin{cases} \lambda_1 M(u_1,u_2) + (1-\lambda_1)C_{1,h}(u_1,u_2), & 0 \le h < h_1; \\ \vdots & \vdots \\ \lambda_i C_{i-1,h}(u_1,u_2) + (1-\lambda_1)C_{i,h}(u_1,u_2), & h_{i-1} \le h < h_i; \\ \vdots & \vdots \\ \lambda_n C_{n-1,h}(u_1,u_2) + (1-\lambda_1)\Pi(u_1,u_2), & h_{n-1} \le h < h_n. \end{cases} \tag{2.2}$$

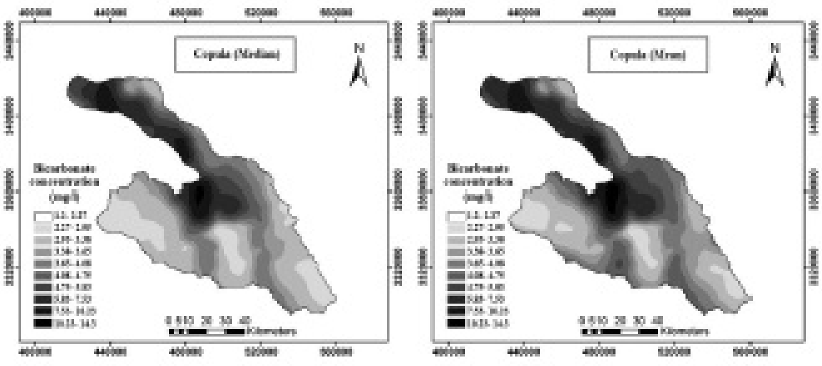

FIGURE 1. zoning map with copula method using mean and median for bicarbonate

In above model λ_i is obtained according to distances of every point from class bounds. Bardossy (2006) got conditional copula to evaluate one point by using n Neighborhood points, and he considered median conditional copula as an estimation of distribution function at unknown point like following model:

$$C^{-1}(U_1|F(x_2),\ldots,F(x_{k+1}))|_{U_1=0.5} \tag{2.3}$$

If F be a distribution function of regional variable, so the estimation of unknown point is inverse F function at the unknown point, it means:

$$Z_{Median}(x_1) = F^{-1}(C_{k+1}^{-1}(0.5|F(x_2),\ldots,F(x_{k+1}))) \tag{2.4}$$

Also, the point estimate of mean value from unknown value can be calculated by use of following Equation:

$$Z_{Mean}(x_1) = \int_0^1 F^{-1}(u).c(u|F(x_2),\ldots,F(x_{k+1}))du \tag{2.5}$$

Aas et al. (2009) presented a new way to transform n-variate copula to $n(n-1)/2$ bivariate copula. This approach provides the possibility of combining the copulas with different families. Two types of structures were used in previous study but in this study we have used a new structure called canonical vine (Aas et al., 2009). The overall structure copula density in canonical vine is as follows:

$$\prod_{k=1}^{n} f(x_k) \prod_{j=1}^{n-1}\prod_{i=1}^{n-j} c_{(j,j+i|1,\ldots j-1)}\{F(x_j|x_1\ldots x_{j-1}),F(x_{j+i}|x_1\ldots x_{j-1})\} \tag{2.6}$$

3. CONCLUSION

Figure 2 show the zoning map with copula method using mean and median for bicarbonate and Table 2, shows the error values of different methods, according to RMSE. The error value was calculated by using the so-called cross validation leave-one-out method. According to Table 1, it can be

Table 1. statistics criteria of the parameters

Parameter	Copula (Median)	Copula (Mean)	Kriging (Simple)	Kriging (Logarithmic)	Kriging (Box-Cox)	IDW
Bicarbonate	1.54	1.73	1.82	1.86	1.81	2.00
Sulfate	6.55	6.61	6.92	6.65	6.93	6.94
Calcium	3.21	3.39	3.58	3.35	3.57	3.66
TDS	929.12	938.21	998.35	934.22	997.41	1055.17

noted that copula has a higher ability than Kriging methods in qualitative zoning of groundwater sources. Also, as you see, the estimation of values in copula method and by use of median has been done with less error than using mean in copula method. One reason for this is that, in estimating by using of median, the value of parameter is considered in the most likely value of copula, but when mean is used, the value of parameter for each copula is considered about 0.5. However, errors of both used methods in all cases were higher than the others, excepting of Calcium estimation with logarithmic transformation in comparison to mean based copula in Kriging method. Accordingly, it can be said that the results of this study indicate copula method has a higher accuracy that other methods such as Kriging and IDW.

References

1. Aas, K., Czado, C., Frigessi, A. and Bakken, H. (2009), Pair-copula constructions of multiple dependence, *Insurance: Mathematics and economics*, **44**(2),182-198.
2. Bardossy, A. (2006), Copula-based geostatistical models for groundwater quality parameters, *Water Resources Research*, **42**(11),1-12.

The 3rd workshop on
Copula and various types of dependencies
25-26, February 2015

NON-LINEAR REGRESSION UNDER PROGRESSIVELY TYPE-II CENSORED ORDER STATISTICS WITH DEPENDENT COMPONENTS

*KARIMI, Z. MADADI, M. AND REZAPOUR, M.

Department of Statistics, Faculty of Mathematics and Computer, Shahid Bahonar University of kerman, Kerman, Iran;
Infozohrehkarimi9055@gmail.com; madadi@uk.ac.ir; mohsenrzp@gmail.com

Abstract. In this paper, we obtain Non-linear regression for PCOS-II arising from dependent units that are jointly distributed according to the Clayton family with an example.

1. Introduction

Let $X_1, \ldots, X_n$ dependent and identical random variables distributed according to an Archimedean copula with completely monotone generator with joint survival function as

$$P(\mathbf{X} > \mathbf{x}) = \psi\left(\sum_{i=1}^{n} \psi^{-1}(\bar{F}(x_i))\right) = \int_0^\infty \prod_{i=1}^{n} G^{\alpha}(\bar{F}(x_i)) dM_\psi(\alpha)\,,$$

where $\mathbf{x} = (x_1, \ldots, x_n)$, $\psi : \mathbb{R}_+ \to [0,1]$ is an n-monotone $(n \geq 2)$ function such that $\psi(0) = 1$ and $\lim_{x\to\infty} \psi(x) = 0$ and $\bar{F} = 1 - F$ is the survival function of X_i, $i = 1, \ldots, n$. Let us further assume that F has density function f and the function G has the first derivative g. Suppose $X^{\mathbf{R}}_{1:m:n}, \ldots, X^{\mathbf{R}}_{r:m:n}$ are the PCOS-II of size m from an dependent sample of size n with a progressive censoring scheme $\mathbf{R} = (R_1, \ldots, R_m)$. Rezapour, obtain the joint probability density function of the first r progressively Type-II Censored

2010 *Mathematics Subject Classification.* Primary 62G02; Secondary 62N01, 62G05.
Key words and phrases. Copula, Archimedean copula, Progressive Type-II Censored Order Statistics.
* Speaker.

Order Statistics as given by

$$f_{\mathbf{X}^{\mathbf{R}}}(x_1,\ldots,x_r)=\int_0^\infty\Big(\prod_{j=1}^{r}\gamma_j\Big)[\prod_{j=1}^{r-1}g(x_j,\alpha)\bar{G}^{R_j}(x_j,\alpha)]g(x_r,\alpha)$$
$$\times\bar{G}^{(\gamma_r-1)}dM_\psi(\alpha)\,,$$

where $\mathbf{X}^{\mathbf{R}}=(X^{\mathbf{R}}_{1:m:n},\ldots,X^{\mathbf{R}}_{r:m:n})$ $\gamma_j=\sum_{v=j}^{m}(R_v+1)$, $\gamma_1=n$, $\bar{G}(x,\alpha)=\exp\Bigg(-\alpha\psi^{-1}\Big(\bar{F}(x)\Big)\Bigg)$, $g(x,\alpha)=\frac{\partial}{\partial x}\bar{G}(x,\alpha)$.
In this paper, we present the Non-linear regression based on PCOS-II for prediction the $(r+s)^{th}$ PCOS-II given r^{th} PCOS-II.

2. Non-linear regression prediction under PCOS-II Arising from clayton family

In this section, we assume $\mathbf{X}^{\mathbf{R}}=(X^{\mathbf{R}}_{1:m:n},\ldots,X^{\mathbf{R}}_{r:m:n})$, the first r PCOS-II arising from units that are jointly distributed according to a clayton Archimedean copula. The generator of Clayton family is $\psi(s)=(1+s)^{\frac{-1}{\theta}}$ for $\theta>0$, therefore, we have $\psi^{-1}(s)=s^{-\theta}-1$. Thus, the joint density function of the $\mathbf{X}^{\mathbf{R}}$ equals

$$f_{\mathbf{X}^{\mathbf{R}}}(x_1,\ldots,x_r)=\int_0^\infty\Big(\prod_{j=1}^{r}\gamma_j\Big)\alpha^r\theta^r\Bigg(\prod_{j=1}^{r}[1-F(x_j)]^{-1-\theta}f(x_j)\Bigg)$$
$$\times\exp\Bigg(-\alpha\Bigg(\sum_{j=1}^{r-1}\Big([1-F(x_j)]^{-\theta}-1\Big)(1+R_j)$$
$$+\gamma_r\Big([1-F(x_r)]^{-\theta}-1\Big)\Bigg)\Bigg)dM_\psi(\alpha)\,. \tag{2.1}$$

Therefore, under the Clayton family assumptions, the marginal density function of $X^{\mathbf{R}}_{r:m:n}$ and the bivariate density function of $(X^{\mathbf{R}}_{r:m:n},X^{\mathbf{R}}_{r+s:m:n})$, $1\le r<r+s\le m$,

$$f_{X^{\mathbf{R}}_{r:m:n}}(x_r)=\frac{c_{r-1}f(x_r)}{(1-F(x_r))^{1+\theta}}\sum_{i=1}^{r}a_i(r)\Bigg(1+\gamma_i\Big((1-F(x_r))^{-\theta}-1\Big)\Bigg)^{-\frac{\theta+1}{\theta}}\,,$$

$$f_{X^{\mathbf{R}}_{r:m:n},X^{\mathbf{R}}_{r+s:m:n}}(x_r,x_{r+s})=\frac{c_{r+s-1}(1+\theta)f(x_{r+s})f(x_r)}{(1-F(x_r))^{1+\theta}(1-F(x_{r+s}))^{1+\theta}}\sum_{j=r+1}^{r+s}\sum_{i=1}^{r}$$
$$\times a_i(r)a_j^{(r)}(r+s)(1+\gamma_j\eta(x_{r+s},x_r;\theta)+\gamma_i\zeta(x_r;\theta))^{\frac{-(2\theta+1)}{\theta}}\,.$$

where

$$\eta(x_{r+s}, x_r; \theta) = \Big((1-F(x_{r+s}))^{-\theta} - (1-F(x_r))^{-\theta}\Big), c_{r-1} = \prod_{j=1}^{r} \gamma_j$$

$$\zeta(x_r; \theta) = \Big((1-F(x_r))^{-\theta} - 1\Big), a_j^{(r)}(r+s) = \prod_{i=r+1, j\neq i}^{r+s,} \frac{1}{\gamma_i - \gamma_j}.$$

To obtain the Non-linear regression we need the estimated parameters. For estimating the parameter θ, we should maximizes the likelihood function. Therefore, the value of θ which maximizes the likelihood function is obtained by solving the equation

$$0 = \frac{\theta^{r-1}(-1)^r \prod_{j=1}^{r} f(x_j)\psi^{(r)}(g(\mathbf{x}, \mathbf{R}; \theta))}{(\prod_{j=1}^{r}(1-F(x_j))^{1+\theta}}\big(r - \theta\sum_{j=1}^{r}\ln(1-F(x_j))$$
$$+\frac{1+r\theta}{1+g(\mathbf{x}, \mathbf{R}; \theta)}(\gamma_r(1-F(x_r))^{-\theta-1}\ln(1-F(x_r)) + \sum_{j=1}^{r-1}(1+R_j)$$
$$\times(1-F(x_j))^{-\theta}\ln(1-F(x_j))\big), \tag{2.2}$$

where $g(\mathbf{x}, \mathbf{R}; \theta) = \gamma_r((1-F(x_r))^{-\theta}-1) + \sum_{j=1}^{r-1}((1-F(x_j))^{-\theta}-1)(1+R_j)$. Non-linear regression for prediction of $X_{r+s:m:n}^{\mathbf{R}}$, based on $X_{r:m:n}^{\mathbf{R}} = x$ can be obtained by $E(X_{r+s}|X_r = x_r)$ which equals

$$D(r, s, x_r, \gamma_r; \theta)\sum_{j=r+1}^{r+s}\sum_{i=1}^{r} a_i(r)a_j^{(r)}(r+s)\Bigg\{\frac{x_r}{\gamma_j}(1+\gamma_i([1-F(x_r)]^{-\theta}$$
$$-1))^{-\frac{1+\theta}{\theta}} + \frac{1}{\gamma_j}\int_{x_r}^{\infty}(\gamma_j[1-F(x_{r+s})]^{-\theta} + B(x_r, \gamma_i, \gamma_j; \theta))^{-(\frac{1+\theta}{\theta})}dx_{r+s}\Bigg\}, \tag{2.3}$$

where

$$D(r, s, x_r, \gamma_r; \theta) = \frac{c_{r+s-1}}{c_{r-1}\sum_{i=1}^{r} a_i(r)(1+\gamma_i\Big([1-F(x_r)]^{-\theta}-1\Big))^{-\frac{1+\theta}{\theta}}},$$
$$B(x_r, \gamma_i, \gamma_j; \theta) = 1 - \gamma_i + (1-F(x_r))^{-\theta}(\gamma_i - \gamma_j).$$

Example 2.1. Let $\mathbf{X} = (x_1, \ldots, x_n)$ has an exponential distribution with distribution function F and density function$f$

Table 1. MLE of parameters θ and λ

θ	λ	$\widehat{\theta}$	$\widehat{\lambda}$
0.5	0.8	0.564	0.78
2	1	1.89	1.06
1	2	1.26	2.18
3	5	2.98	4.48
6	8	6.4	7.56

Non-linear regression for prediction of $X^{\mathbf{R}}_{r+s:m:n}$, based on $X^{\mathbf{R}}_{r:m:n} = x$ can be obtained by $E(X_{r+s}|X_r = x_r)$ which equals

$$D(r,s,x_r,\gamma_r;\widehat{\theta},\widehat{\lambda}) \sum_{j=r+1}^{r+s} \sum_{i=1}^{r} a_i(r) a_j^{(r)} (r+s)(\frac{x_r}{\gamma_j}[1+\gamma_i(\exp(\widehat{\lambda}\widehat{\theta}x_r)-1)]^{-(\frac{1+\widehat{\theta}}{\widehat{\theta}})}$$
$$+\frac{1}{\widehat{\lambda}\widehat{\theta}\gamma_j}\int_0^{\frac{-1}{1+\gamma_i[\exp(\widehat{\lambda}\widehat{\theta}x_r)-1]}})(\frac{1}{z} - B(r,s,x_r,\gamma_r;\widehat{\theta},\widehat{\lambda}))^{-1} z^{\frac{1}{\widehat{\theta}}-1} dz),$$

where

$$D(r,s,x_r,\gamma_r;\widehat{\theta},\widehat{\lambda}) = \frac{c_{r+s-1}}{c_{r-1}\sum_{i=1}^{r} a_i(r)\Big(1+\gamma_i[\exp(\widehat{\lambda}\widehat{\theta}x_r)-1]^{\frac{-(1+\widehat{\theta})}{\widehat{\theta}}}\Big)},$$

and

$$B(r,s,x_r,\gamma_r;\widehat{\theta},\widehat{\lambda}) = 1-\gamma_i + (\gamma_i - \gamma_j)\exp(\widehat{\lambda}\widehat{\theta}x_r).$$

References

1. Balakrishnan, N. (2007), Progressive censoring methodology: an appraisal (with Discussions), *Test*, **16**, 211-296.
2. Balakrishnan, N. and Aggarwala, R. (2000), Progressive Censoring: *Theory, Methods, and Applications*, Birkhauser, Boston.
3. Joe, H. (1997), *Multivariate models and dependenc concept*, Chapman and Hall, London.
4. Mcneil, A.J. and Neslehova, J. (2009), Multivariate Archimedean copulas, d-monotone functions and L1norm symmetric distributions, *Ann. Stat.*, **37**, 3059-3097.
5. Nelsen, R.B. (2006), An introduction to Copulas, *Springer*, New York.
6. Rezapour, M., Alamatsaz, M. H. and Balakrishnan, N. (2013). On properties of dependent general progressively Type-II censored order statistics, *Metrika*, **76**, 909-917 .
7. Rezapour, M., Alamatsaz, M. H., Balakrishnan, N. and Cramer, E. (2013). On properties of progressively Type-II censored order statistics arising from dependent and non-identical random variables, *Statistical methodology*, **10**, 58-71.

The 3rd workshop on
Copula and various types of dependencies
25-26, February 2015

CONSTRUCTION A FAMILY OF NONSEPARABLE SPATIO-TEMPORAL COVARIANCES BY USING COPULA FUNCTIONS

MEHDI OMIDI AND MOHSEN MOHAMMADZADEH*

Department of Statistics, Tarbiat Modares University, P. O. Box 14115-137, Tehran, Iran;
mohsen_m@modares.ac.ir

Abstract. Statistical analysis of natural phenomena with spatial and temporal correlations requires to specify their correlation structure via a covariance function. A separable spatio-temporal covariance function is usually used for the ease of application. Nonetheless, the separability of the spatio-temporal covariance function can be unrealistic in many settings, where it is required to use a non-separable spatio-temporal covariance function. In this paper, a structural copula function is applied to construct a family of non-separable spatio-temporal covariance function. Next, a modified genetic algorithm is applied to explore the spatio-temporal correlation structure of Ozone data in Tehran, Iran.

1. Introduction

The space-time model for analyzing the spatial data observed over time has received more attentions in recent years. In this line, many researchers have introduced various classes of valid nonseparable spatio-temporal covariance functions (see e.g. [1], [2] and [8]).

One of the drawbacks of using such a nonseparable space-time covariance functions, however, is the availability of many parameters.

2010 *Mathematics Subject Classification.* Primary 62H11; Secondary 91B72.
Key words and phrases. Copula Function, Spatio-temporal Covariance, Genetic Algorithm.
* Speaker.

This shortcoming encourages experts to use very practical and robust optimisation methods to estimate parameters. Among many capable computational methods, the Genetic Algorithm (GA) suggests a flexible and broader range of solutions for finding the extremum values.

In the present paper, a copula function is used to construct some nonseparable spatio-temporal covariance functions. Moreover, to overcome the problems of the parameter estimation, we modified Binary GA proposed by [4] in a more sensible way. All proposed models and techniques have been applied to analyse the Ozone concentration collected over the entire year 2012 in Tehran, the capital city of Iran.

2. Background Concepts

For two continuous random variables X and Y with distribution functions $F_X(x)$ and $G_Y(y)$ respectively, and joint distribution function $H(\cdot,\cdot)$, by Sklar's Theorem ([9]) a unique copula function $K(\cdot,\cdot)$ exists such that

$$H(x,y) = K(F_X(x), G_Y(y)), \ (x,y) \in \bar{\Re}^2.$$

Suppose $\varphi(t) : [0,1] \to [0,\infty]$ is a decreasing function satisfying $\varphi(1) = 0$ with inverse $\varphi^{-1}(\cdot)$, then

$$K_\varphi(F_X(x), G_Y(y)) = \varphi^{-1}(\varphi(F_X(x)) + \varphi(G_Y(y))) \tag{2.1}$$

is a copula function if and only if $\varphi(\cdot)$ is convex ([7]). A class of copula functions made by Eq. (2.1) is known as Archimedean family, with a generating function $\varphi(\cdot)$.

For the theoretical results in subsequent sections, some particular functions need to be defined that will be used to construct a spatial covariance function.

Definition 2.1. A function $\phi(\cdot)$ is called a Completely Monotone (CM) function on an interval I, if all of its derivatives exist on I and satisfies in

$$(-1)^n \phi^{(n)}(t) \geq 0, \ n = 0, 1, \ldots$$

where $\phi^{(n)}$ denotes the n'th derivative of $\phi(\cdot)$.

Based on Bernstein's Theorem, $\phi(\cdot)$ is CM if and only if it is the Laplace transformation of positive measure M$(\cdot)$. Then for all $t > 0$, the following exists

$$\phi(t) = \int_0^\infty e^{-rt} dM(r) \tag{2.2}$$

TABLE 1. Some Archimedean copula and their corresponding stationary spatial covariance functions.

Copula	Generator	Parameter space	Covariance function
Ali-Mikhail-Haq	$\ln \frac{1-\theta(1-t)}{t}$	$[-1,1]$	$\frac{1-\theta}{\exp(\lvert\lvert h\rvert\rvert^{\alpha})-\theta}$
Joe	$-\ln(1-(1-t)^{\theta})$	$[1,\infty)$	$1-(1-e^{-\lvert\lvert h\rvert\rvert^{\alpha}})^{\frac{1}{\theta}}$
Frank	$-\ln\{\frac{\exp(-\theta t)-1}{\exp(-\theta)-1}\}$	$\Re-\{0\}$	$-\frac{1}{\theta}\ln\{(e^{-\theta}-1)e^{-\lvert\lvert h\rvert\rvert^{\alpha}}+1\}$
Gumbel-Barnett	$\ln(1-\theta\ln t)$	$[0,1)$	$\exp\{\frac{1-e^{\lvert\lvert h\rvert\rvert^{\alpha}}}{\theta}\}$
Gumbel-Hougaard	$(-\ln t)^{\theta}$	$[1,\infty)$	$\exp(-\lvert\lvert h\rvert\rvert^{\alpha})$

A nonnegative function $\psi(t)$, $t>0$ is called Bernstein function if its derivative is CM. Some important properties of CM and Bernstein functions are as follows.

If $\phi_1(t)$ and $\phi_2(t)$ are CM and $\psi_1(t)$ and $\psi_2(t)$ are Bernstein functions, then

(i) $\phi_1(t)+\phi_2(t)$ and $\phi_1(t)\phi_2(t)$ are CM.
(ii) $\phi_1(\psi_1(t))$ is CM.
(iii) $\psi_2(\psi_1(t))$ is Bernstein.

[5] showed that the inverse of a generating function of any Archimedean copula family is a Laplace transformation. Then, the inverse of generating functions of the Archimedean copulas are CM. Therefore, for any $t>0$ and $\rho\in[0,1]$ the function t^{ρ} is Bernstein. By (ii) for any generating function $\varphi(\cdot)$ and $\varphi(0)=\infty$, $\varphi^{-1}(t^{\rho})$ is CM. Thus, according to [6], for any $\alpha=2\rho$, $C(\lvert\lvert h\rvert\rvert)=\varphi^{-1}(\lvert\lvert h\rvert\rvert^{\alpha}), h\in\Re^d, d\geq 1$ is a stationary covariance function.

Table 1 shows some generators of the Archimedean copula family presented in [7] along with their corresponding stationary covariance functions.

3. Construction of Spatio-temporal Covariance Function

A spatio-temporal data is modeled by a random field $\{Z(s,t),(s,t)\in D\times T\}$, where $D\subset\Re^d$, $T\subset\Re$. $Z(s,t)$ is a real-valued stochastic process at the spatial location s, and time t.

If the mean function $\mu(s,t)$ is constant for all (s,t) and covariance function $C(s,s',t,t')$ depends only on spatial lag, $h_s=s-s'$ and temporal lag $h_t=t-t'$, then the random field is called second-order stationary. In this case, the spatio-temporal covariance is denoted by $C(h_s,h_t)$.

[2] showed that if $\phi(\cdot)$ is CM and $\psi(\cdot)$ is Bernstein with $\psi(0) = 1$ and $\psi(\infty) = \infty$, then

$$C_{s,t}(h_s, h_t) = \frac{\sigma^2}{\psi(|h_t|^2)^{d/2}}\phi(\frac{||h_s||^2}{\psi(|h_t|^2)}) \tag{3.1}$$

is a valid stationary spatio-temporal covariance function in $\Re^d \times \Re$.

Example 3.1. Let $\phi(t) = \exp(-\beta t^{r_1})$ and $s(t) = (at^\alpha + 1)^{\frac{-1}{\theta}}$, then

$$Model\, 1 : C_{s,t}(h_s, h_t) = \sigma^2(a|h_t|^\alpha + 1)^{\frac{-d}{2\theta}} \exp\{-\beta[b||h_s||^2(a|h_t|^\alpha + 1)^{\frac{-1}{\theta}}]^{r_1}\}$$

for $a > 0$, $b > 0$, $0 < r_1 \leq 1$, $\alpha \in (0, 2]$, $\beta > 0$, $\theta \geq 1$ is a valid stationary spatio-temporal covariance function in $\Re^d \times \Re$.

[5] showed that for any generator of the Archimedean copula, $\varphi_\theta(t)$, if $\theta_1 \geq \theta_2$ and $\varphi_\theta(0) = \infty$ then $\psi(t) = \varphi_{\theta_2}(\varphi^{-1}_{\theta_1}(t))$ is a Bernstein function such that $\psi(0) = 0$ and $\psi(\infty) = \infty$. Therefore, if $\phi_{\theta'}(\cdot)$ is CM and $\varphi_\theta(t)$ is an Archimedean copula generator, then $\phi_{\theta'}(\varphi_{\theta_2}(\varphi^{-1}_{\theta_1}(||h||^\alpha)))$ and $\varphi^{-1}_{\theta'}(\varphi_{\theta_2}(\varphi^{-1}_{\theta_1}(||h||^\alpha)))$ for $\theta_1 \geq \theta_2$ and $\alpha \in [0, 2]$ are stationary spatial covariance function in $\Re^d$.

[3] showed that if $C^r_s(h_s)$ is a purely spatial covariance in $\Re^d$ and $C^r_t(h_t)$ is a purely temporal covariance in $\Re$, then

$$C_{s,t}(h_s, h_t) = \int_0^\infty C^r_s(h_s)C^r_t(h_t)d\mu(r) \tag{3.2}$$

is a stationary spatio-temporal covariance function in $\Re^d \times \Re$, where $\mu(\cdot)$ is a finite positive measure.

Theorem 3.2. *Let φ, $\varphi_1$and φ_2 be generators of Archimedean copulas such that $\varphi(0) = \varphi_1(0) = \varphi_2(0) = \infty$, then*

$$C_{s,t}(h_s, h_t) = \varphi^{-1}_\theta[\varphi_{1_{\theta_2}}(\varphi^{-1}_{1_{\theta_1}}(a||h_s||^\alpha)) + \varphi_{2_{\theta_4}}(\varphi^{-1}_{2_{\theta_3}}(b|h_t|^\beta))] \tag{3.3}$$

for $\alpha, \beta \in (0, 2]$, $\theta_1 \geq \theta_2$, $\theta_3 \geq \theta_4$ and a, $b > 0$, is a family of stationary spatio-temporal covariance functions in $\Re^d \times \Re$.

Example 3.3. Let φ be the generator of Joe copula, φ_1 generator of Gumbel-Hougaard and φ_2 generator of Clayton copula, then

$$Model\, 2 : C_{s,t}(h_s, h_t) = \sigma^2[1 - (1 - \exp\{-a||h_s||^\alpha - (1 + b|h_t|^\beta)^{r_1} + 1\})^{\frac{1}{\theta}}]$$

for $0 < r_1 \leq 1$, $\alpha, \beta \in (0, 2]$, $\theta > 1$ and a, $b > 0$, is a stationary spatio-temporal covariance function in $\Re^d \times \Re$.

Example 3.4. Let φ be the generator of Gumbel-Hougaard and φ_1 and φ_2 be the generators of Clayton copula, then

$$Model\, 3 : C_{s,t}(h_s, h_t) = \sigma^2 \exp\{-[(1 + a||h_s||^\alpha)^{r_1} + (1 + b|h_t|^\beta)^{r_2} - 2)]^{\frac{1}{\theta}}\}$$

Table 2. Parameter estimates of the models, the best final cost and AIC criterions.

	Model			
Parameter	1	2	3	4
υ_1	5.917	0.042	0.023	1.275
υ_2	2.740	3.750	0.767	7.595
α	0.998	0.067	0.034	0.019
β	2.729	0.893	1.635	0.116
θ	3.650	1.256	1.048	0.237
r_1	0.890	0.534	0.980	0.359
r_2	–	–	0.493	0.116
σ^2	0.323	0.376	0.120	0.126
-Best final cost	995.042	997.175	996.928	994.755
$-AIC$	1968.084	1972.350	1969.856	1965.510

for $0 < r_1 \leq 1$, $0 < r_2 \leq 1$, $\alpha, \beta \in (0, 2]$, $\theta > 1$ and a, $b > 0$, is a stationary spatio-temporal covariance function in $\Re^d \times \Re$.

Example 3.5. Let φ, φ_1 and φ_2 be generators of Clayton copula, then

$$Model\,4 : C_{s,t}(h_s, h_t) = \sigma^2\{(1 + a||h_s||^{\alpha})^{r_1} + (1 + b|h_t|^{\beta})^{r_2}) - 1\}^{\frac{-1}{\theta}}$$

for $0 < r_1 \leq 1$, $0 < r_2 \leq 1$, $\alpha, \beta \in (0, 2]$, $\theta > 1$ and a, $b > 0$, is a stationary spatio-temporal covariance function in $\Re^d \times \Re$.

4. Application

In this section, the models 1 to 4 proposed in the previous section are applied to determine the spatio-temporal correlation structure of ozone data in Tehran, Iran. This megacity with a population of over 10 million, is suffering from high air pollution. The weekly average of ozone concentration in 2012 were observed at 9 different pollution monitoring stations, namely Sorkhe Hesar, Golbarg, Aghdasiyeh, Masoodiyeh, Ray, Geophysics, Azadi, Ponak and Zarpark.

Exploratory data analysis revealed non-normality of the ozone data, but the Shapiro-wilk test with $p\_value = 0.146$ was significant for the transformed data by Box-Cox transformation $\frac{x^{\lambda}-1}{\lambda}$, with $\lambda = 0.2181$. As a result, the transformed data outcomes are taken into account as a Gaussian spatio-temporal random field $Z(s, t)$.

Regarding the scale parameters of space and time, we assumed that $a = (\frac{1}{\upsilon_1})^{\alpha}$, $\upsilon_1 > 0$ and $\alpha \in [0, 2]$ for all models, $b = (\frac{1}{\upsilon_2})^2$ in model 1, $b = (\frac{1}{\upsilon_2})^{\beta}$, $\upsilon_2 > 0$ and $\beta \in [0, 2]$ for models 2, 3 and 4.

For estimate parameters of all models Genetice algorithm was used and the parameter estimates, the best final cost and AIC criterion are shown in Table 2. Because of the minimum values of AIC and the best final cost, Model 2 is relatively better among the others for structural modelling of spatio-temporal correlation of ozone data in Tehran.

5. Conclusion

To analyse the spatio-temporal data, it is necessary to use a valid non-separable spatio-temporal covariance function. In the present paper, copula function is used to construct spatial and spatio-temporal covariance functions. For the estimation of the model parameters, the GA was used with a new selection strategy. Then, to ensure the accuracy and precision of the estimations, the best cost of likelihood function was found by repeating the GA in ten times. The results demonstrate that GA is an enormously powerful and successful algorithm in the estimation of the spatio-temporal covariance model.

References

1. Cressie, N. and Huang, H.C. (1999), Classes of Non-Separable, Spatio-Temporal Stationary Covariance Functions. Journal of the American Statistical Association, 94, 1330-1340.
2. Gneiting, T., (2002), Nonseparable, Stationary Covariance Functions for Space-Time Data. Journal of the American Statistical Association, 97, 590-600.
3. Gneiting, T., Genton, M. G. and Guttorp, P., (2007), Geostatistical Space-time Models, Stationarity, Separability and Full Symmetry, In Finkenstadt, B., Held, L. and Isham, V. (eds.), Statistical Methods for Spatio-Temporal Systems. Chapman and Hall. 151-175.
4. Haupt, R. L. and Haupt, S. E., (2004), Practical Genetic Algorithms. Second Edition, John Wiley and Sons, New York.
5. Joe, H., (1997), Multivariate Models and Dependence Concepts, Chapman and Hall.
6. Ma, C., (2005) Linear Combinations of Space-Time Covariance Functions and Variograms. IEEE Transactions on Signal Processing, 53(3), 857-864.
7. Nelsen, R. B., An Introduction to Copulas. Second Edition, New York, Springer, (2006).
8. Rodrigues, A. and Diggle, P., (2010), A Class of Convolution-based Models for Spatio-temporal Processes with Non-separable Covariance Structure. Scandinavian Journal of Statistics, 37(4), 553-567.
9. Sklar, A., (1959), Fonctions de Repartition a n Dimensions et Leurs Marges. Publications de l'Institut de Statistique de l'Universite de Paris 8, 229-231.

The 3rd workshop on
Copula and various types of dependencies
25-26, February 2015

A COPULA APPROACH TO BIVARIATE CREDIBLE CONFIDENCE INTERVAL FOR FREQUENCY AND SEVERITY OF BONUS–MALUS SYSTEMS

PARVASIDEH, L.*, PAYANDEH NAJAFABADI, A. T. AND BAHRAMI SAMANI, E.

Department of Mathematical Sciences, Shahid Beheshti University, G.C. Evin, 1983963113, Tehran, Iran.
L_parvasideh@yahoo.com; amirtpayandeh@sbu.ac.ir; e_bahrami@sbu.ac.ir

Abstract. The most of advanced Bonus-Malus Systems works based upon both frequency and severity of reported claims. This article utilizes copula idea to develop a bivariate credible confidence interval for frequency and severity of a given Bonus–Malus System.

1. Introduction

In many countries insurers use Bonus-Malus system, say BMS, in order to provide fair premium amounts based on policyholders claim experience. Such system penalizes insured drivers who claim at least one accident (malus) and rewards claim-free drivers (bonuses). In practice, a BMS consists of a finite number of levels, numbered from 1 to s, as policyholders risk classification. In fact, the policyholder who has the smallest risk-tendency stands in the first level and pays the smallest premium. In the same manner, the policyholder who has the largest risk-tendency stands the last level and pays the largest premium.

Martin & Lof (1973) are the first authors, who study theory of BMS. Up to now, several books and papers have been written about BMS. One of the complete sources of this, we can mention to Lemaire (1995) and Dionne (2001). Basic objection of classic BMSs is that transmittal rules

2010 *Mathematics Subject Classification.* Primary 47A55; Secondary 39B52, 34K20, 39B82.

Key words and phrases. Clayton copula; Credible confidence intervals; Bayesian analysis; Geometric quantiles; Optimal Bonus–Malus Systems .

* Speaker.

and their premium only is based on number of losses without considering size of reported claims. This approach may be a policyholder who had accident with a small size of loss is penalized like a policyholder who had accident with big size of losses. For this problem, actuaries proposed the optimal BMS designed not only on the number of losses but also on the size of losses [1].

With using definition of his system, we find the joint distribution of number (frequency) and size (severity) of claims by using copula functions. We utilize independence and clayton copula in this direction. Then, we construct simultaneous credible confidence interval for both of number and size of reported claims. To find such simultaneous credible confidence interval, we utilize geometric quanties.
Copula approach provides a practical method to take into account dependence between random variables from their marginal distributions. Consider any continuous random variables $X_1, \cdots, X_d$ with corresponding distribution functions $F_1, \cdots, F_d$. Joint distribution function can be restated as

$$F(x_1, \cdots, x_d) = C_\theta[F_1(x_1), \cdots, F_d(x_d)], \tag{1.1}$$

where C_θ is a copula function with parameter θ.. If $(X_1, \cdots, X_d) \in R^d$ has a continuous multivariate distribution with $F(x_1, \cdots, x_d) = P(X_1 \leq x_1, \cdots, X_d \leq x_d)$, then by Sklar's theorem there is a unique copula function $C : [0,1]^d \longrightarrow [0,1]$ of F such that,

$$F(x_1, \cdots, x_d) = C(F_1(x_1), \cdots, F_d(x_d)). \tag{1.2}$$

One of the important family of copula, is Archimedean copulas that construct based on generator ϕ [5]. This paper employs the following Clayton copula (a member of the Archimedean copulas) to develop the desired simultaneous credible confidence interval.

$$C(u, v) = (u^{-\theta} + v^{-\theta} - 1)^{-1/\theta}, \theta \in (0, \infty) \tag{1.3}$$

This paper structured as the following. Section 2 represents main results. Practical application of our findings have been provided in Section 3.

2. Main results

The following two theorems provide posterior distribution of parameters of frequency and severity.

Theorem 2.1. *Suppose $X_1, \cdots, X_t$ stands for the number of reported accidents in t years. Moreover, suppose that $X_1, \cdots, X_t$ are i.i.d random variable has been distributed according to a Poisson distribution (with mean λ) whenever parameter λ has been given. Also suppose that λ has prior distribution $Gamma(a, b)$. Then posterior distribution $\lambda | X_1, \cdots, X_t$ is the $Gamma(a + K, b + t)$ distribution which $K = \sum_{i=1}^{t} x_i$.*

Proof. For proof use from bayes rule.

$$\pi(\lambda \mid x) = \frac{f(x \mid \lambda) f(\lambda)}{\int f(x \mid \lambda) f(\lambda)\, d\lambda}$$

First calculate denominator.

$$\begin{aligned} A = \int f(x \mid \lambda) f(\lambda)\, d\lambda &= \int \frac{e^{-\lambda t} \lambda^{\sum_{i=1}^t x_i}}{\prod_{i=1}^t x_i!} \frac{\lambda^{a-1} b^a e^{-b\lambda}}{\Gamma(a)}\, d\lambda \\ &= \frac{b^a}{\prod_{i=1}^t x_i! \Gamma(a)} \frac{\Gamma(\sum_{i=1}^t x_i + a)}{(t+b)^{\sum_{i=1}^t x_i + a}} \end{aligned}$$

Now, we have

$$\begin{aligned} \pi(\lambda \mid x) &= \frac{e^{-\lambda t} \lambda^{\sum_{i=1}^t x_i}}{\prod_{i=1}^t x_i!} \frac{\lambda^{a-1} b^a e^{-b\lambda}}{\Gamma(a)} \times A^{-1} \\ &= \frac{e^{-\lambda(t+b)} (t+b)^{\sum_{i=1}^t x_i + a} \lambda^{\sum_{i=1}^t x_i + a - 1}}{\Gamma(\sum_{i=1}^t x_i + a)} \end{aligned}$$

□

Theorem 2.2. *Suppose $Y_1, \cdots, Y_t$ stands for the size of reported accidents in t years. Moreover, suppose that $Y_1, \cdots, Y_t$ are i.i.d random variable has been distributed according to an exponential distribution (with mean β) whenever parameter β has been given. Also suppose that λ has prior distribution $IGamma(s, m)$. Then posterior distribution $\beta | Y_1, \cdots, Y_t$ is the $IGamma(s + t, m + L)$ distribution which $L = \sum_{i=1}^t y_i$.*

Proof. Like previous theorem, to calculate denominator we have:

$$\begin{aligned} B = \int f(y \mid \beta) f(\beta)\, d\beta &= \int (\frac{1}{\beta})^t e^{-\frac{\sum_{i=1}^t y_i}{\beta}} \frac{\frac{1}{m} e^{-m/\beta}}{(\frac{\beta}{m})^{s+1} \Gamma(s)}\, d\beta \\ &= \frac{m^{s+1}}{m\Gamma(s)} \frac{(\sum_{i=1}^t y_i + m)\Gamma(s+t)}{(\sum_{i=1}^t y_i + m)^{s+t}} \end{aligned}$$

Now we have

$$\begin{aligned} \pi(\beta \mid y) &= (\frac{1}{\beta})^t e^{-\frac{\sum_{i=1}^t y_i}{\beta}} \frac{\frac{1}{m} e^{-m/\beta}}{(\frac{\beta}{m})^{s+1} \Gamma(s)} \times B^{-1} \\ &= \frac{\frac{1}{\sum_{i=1}^t y_i + m} e^{-\frac{\sum_{i=1}^t y_i + m}{\beta}}}{(\frac{\beta}{\sum_{i=1}^t y_i + m})^{s+t} \Gamma(s+t)} \end{aligned}$$

□

To construct simultaneous credible confidence intervals, first we connect two posterior distribution by a Clayton copula. Then by utilize geometric quantiles (Chaudhuri 1996), obtain this intervals [3]. In this way we define $\mathbf{u}$ which $u = 2\alpha - 1$ and $\alpha \in (0, 1)$ such that $B^{(d)} = \{\mathbf{u} | \mathbf{u} \in R^d, |\mathbf{u}| < 1\}$.

Then the geometric quantile $\hat{\mathbf{Q}}_{\mathbf{n}}(\mathbf{u})$ corresponding to $\mathbf{u}$ and based on d-dimensional data points $\mathbf{X_1}, \ldots, \mathbf{X_n}$ is defined as

$$\hat{\mathbf{Q}}_{\mathbf{n}}(\mathbf{u}) = arg \min_{\mathbf{Q} \in \mathbb{R}^d} \sum_{i=1}^{n} \Phi(\mathbf{u}, \mathbf{X_i} - \mathbf{Q}) \tag{2.1}$$

We let $d = 2$ and generate random variable from joint distribution with Clayton copula. Now we have n point and purpose is calculate αth quantile of this points. To solve the equation can use iterative methods like the Newton-Raphson-type method. One needs an initial approximation of $\hat{\mathbf{Q}}_{\mathbf{n}}(\mathbf{u})$ to start the iteration. Such initial approximation can be the vector of medians of $\mathbf{X}_i$. In this way simultaneous credible confidence intervals is corresponding quantiles of the either posterior distributions.

3. A simulation study

Suppose 10 reported claims in a Bonus-Malus system are available. Moreover suppose that both prior distribution are $\lambda \sim Gamma(0.3, 1/3)$ and $\beta \sim IGamma(2.5, 1/5)$. Now using result of pervious section, we develop bivariate credible confidence interval for frequency and severity of such Bonus–Malus System.

Table 1: Credible confidence interval for frequency and severity under independent copula.

Copula	90%	95%
Independent copula	$(0.07, 0.74) \times (2.52, 7.76)$	$(0.057, 0.83) \times (2.35, 8.58)$

From this result, we conclude with 90% confidence, a person will be had 0.07 to 0.74 acccident and will be had 2.52 to 7.76 unit of money severity of claims in next year.

Table 2: Credible confidence interval for frequency and severity under Clayton copula.

Copula	70%	60%
Clayton copula	$(0.29, 5.49) \times (2.88, 9.78)$	$(0.29, 1.52) \times (3.07, 5.93)$

From this result, we conclude with 70% confidence, a person will be had 0.29 to 5.49 acccident and will be had 2.88 to 9.78 unit of money severity of claims in next year.
Note that since we have definition $\mathbf{B}^{(d)}$, so in a seconde way, Unfortunately confidence level is less and interval is larger than first way.

4. Conclusion and suggestion

This article employs a Clayton copula to provide a credible confidence interval for frequency and severity of a Bonus–Malus System. Using such credible confidence interval, one may improve actuarial inference about a Bonus–Malus System. Using other copula such actuarial inference may be improved.

References

[1] Dudzinski, M. ,Furmanczyk, K. and Kocinski, M. (2013), Bayesian confidence intervals for the number and the size of losses in the optimal Bonus-Malus system. *Quantitative Methods In Economics*, **1**(14) , 93-104.

[2] Lemaire, J. (1995), *Bonus-Malus System in automobile insurance*, Kluwer Academic Publisher.

[3] Chaudhuri, P. (1996), On a geometric notion of quantiles for multivariate data. *Journal of the American Statistical Association*, **91**(434), 862-872.

[4] Dionne, G. ,Gouriroux, C. and Vanasse, C. (2001), Testing for Evidence of Adverse Selection in the Automobile Insurance Market. *Journal of Political Economy*, **2** (109) , 444-453.

[5] Klugman, S.A. ,Panjer, H.H. and Willmot, G.E. (2008), *Loss models*, Wiley Series.

The 3rd workshop on
Copula and various types of dependencies
25-26, February 2015

THE MOST COMMON GOODNESS OF FIT TESTS FOR COPULA MODELS

POURAHMADI, M. [1] AND BEHBASH, Z.[2*]

[1] *Department of Biostatistics, Student Research Committee, Shahid beheshti University of Medical Sciences, Tehran, Iran*
m.pourahmadi66@gmail.com

[2] *Department of Statistics, Faculty of Mathematical and Statistics, University of Shahid Chamran, Ahvaz, Iran*
z.behbash@yahoo.com

Abstract. Copulae is one of the main ways of modelling dependence. Many proposals have been made recently for goodness of fit testing of copula models. In this paper we propose and analysis the several most common methods of goodness of fit test that use for copula selection. We eventually apply these methods to select a suitable copula of the two variables associated with the Iran's financial data: gross domestic production, oil income index.

Introduction

The definition of a d-dimensional copula is a multivariate distribution C, with uniform margins $U(0,1)$. Sklar (1959)'s theorem states that every multivariate distribution F with margins $F_1, F_2, \ldots, F_d$ can be written as

$$F(x_1, \ldots, x_d) = C(F_1(x_1), F_2(x_2), \ldots, F_d(x_d))$$

for some copula C. For more information about analysis of copulae, see Joe (1997) or Nelsen (1999).
Goodness of fit testing for copulae recently emerged as a challenging inferential problem and some approaches have been proposed. The limitation of the copula approach is the lack of a recommended way of checking whether the dependency structure of a data set is appropriately modeled by a chosen

Key words and phrases. Blanket tests, Clarke-test, Kendall test, Vuong test.
* Speaker.

family of copulae. According to Genest et al. (2009) copula goodness of fit tests in the literature can be divided into three groups:
(*i*) Procedures for testing specific parametric copula families such as the Normal or Clayton families.
(*ii*) General tests for any copula family but which involve some kind of parameter tuning or other strategic choices of smoothing parameter, weight function or kernel.
(*iii*) So-called blanket tests which are applicable to all copula structures and do not involve any preliminary strategic choices as in (*ii*).
Here we will concentrate on the last group and, in particular, on two blanket tests based on the empirical copula process and on Kendall's transformation and on Vuong and Clarke-teste, since we are interested in general procedures without any limitations in its use. These tests are often used in statistical hypothesis testing, in tests based on the empirical copula process and on Kendall's transformation we test the hypothesis if a chosen copula fits the underlying copula of the data.

$$H_0 : C \in C_0 = \{C_\theta : \ \theta \in \Theta\} \quad vs \quad H_1 : C \notin C_0 \qquad (1)$$

Where C_0 is the set of copulas and Θ is the parameter space.
Two tests based on empirical copula process
Let $\mathbf{v}_1 = (\mathrm{v}_{11}, \ldots, \mathrm{v}_{n1}), \ldots, \mathbf{v}_d = (\mathrm{v}_{1d}, \ldots, \mathrm{v}_{nd})$ be $U(0,1)$ distributed random samples with copula C. Suppose it is desired to test the null hypothesis (1). Naturally, we want to compare the distance between the empirical copula C_n, i.e.

$$C_n(\mathrm{u}) = \frac{1}{n}\sum_{i=1}^{n} 1\{\mathrm{v}_{i1} \leq \mathrm{u}_1, \ldots, \mathrm{v}_{id} \leq \mathrm{u}_d\} \quad , \quad \mathbf{u} = (\mathrm{u}_1, \ldots, \mathrm{u}_d) \in [0,1]^d$$

and the parametric estimate $C_{\hat{\theta}}$, where $\hat{\theta}$ is an estimate of the unknown parameter θ. Based on this concept, Genest and Remillard [2008] established the empirical copula process

$$\mathbb{C}_n = \sqrt{n}(C_n - C_{\hat{\theta}})$$

which measures the distance $(C_n - C_{\hat{\theta}})$ with a scale $\sqrt{n}$.
Genest and Remillard [2008] considered rank-based versions of the familiar Carmer von Mises and Kolmogorov Smirnov statistics in combination with

$$S_n^{\mathbb{C}_n} = \int_{[0,1]^d} \mathbb{C}_n^2(\mathrm{u}) dC_n(\mathrm{u}) \quad \text{and} \quad T_n^{\mathbb{C}_n} = \sup_{\mathrm{u} \in [0,1]^d} |\mathbb{C}_n(\mathrm{u})|$$

We call GOF tests based on these statistics tests based on the empirical copula process.
Two tests based on Kendall's transform
These tests are explored by Genest and Rivest [1993], and Wang and Wells [2000]. Let $\mathbf{X} = (\mathrm{X}_1, \ldots, \mathrm{X}_d)$ be a continuous d-variate random vector with distribution function F, margins $F_1, \ldots, F_d$ and unique underlying copula

C. Let $\text{U}_i = F_i(\text{X}_i)$ for $i = 1, \ldots, d$, then the joint distribution of $\mathbf{U} = (\text{U}_1, \ldots, \text{U}_d)$ is C. Suppose we are interested in Hypothesis

$$H_0 : C \in C_0 = \{C_\theta : \ \theta \in \Theta\}$$

Now under H_0, the vector $\mathbf{U}$ is distributed as C_θ for some $\theta \in \Theta$. Let K_θ denote the Kendall distribution function of C_θ, and K_n denote the corresponding empirical Kendall distribution function which is an estimator of K_θ. Hence, $C_\theta(\text{U})$ has distribution K_θ. Through the Kendall process

$$\mathbb{K}_n(t) = \sqrt{n}(k_n(t) - k_{\theta_n}(t))$$

one can test

$$H_0^{''} : k \in K_0 = \{k_\theta : \ \theta \in \Theta\}$$

More discussion about this limitation can be found in Wang and Wells [2000] and Genest et al. [2007]. The specific test statistics for this GOF test are given by

$$S_n^{\mathbb{K}_n} = \int_0^1 |\mathbb{K}_n(t)|^2 dk_{\theta_n}(t)$$
$$T_n^{\mathbb{K}_n} = \sup_{0 \leq t \leq 1} |\mathbb{K}_n(t)|$$

Vuong and Clarke-test

The Vuong and the Clarke-test (Vuong [1989], Clarke [2007]) are tests to compare two models, which are not necessarily nested. Both are based on the likelihood or rather on their likelihood ratio and the Kullback-Leibner information criterion (KLIC). The KLIC between the true copula C_0 and an alternative copula C_1 can be rewritten as

$$\text{KLIC}(C_0, C_1) = \int_{[0,1]^2} c_0(u, v) \log \left[\frac{c_0(u, v)}{c_1(u, v)} \right] dudv$$

where c_0 and c_1 denote the copula densities corresponding to copulas C_0 and C_1, respectively. The model with the minimum KLIC, i.e. the smallest distance, is the best one.

To compare two models and on the basis of the likelihood ratio from above, Vuong [1989] defined and calculated the following statistics: (see also Erhardt [2006]).

As we see, If model 1 is better than model 2, its KLIC statistic is smaller and the following inequality holds

$$\text{KLIC}_1(u, v) < \text{KLIC}_2(u, v)$$

This expression reduces to

$$E\left[\log\left(\frac{c_1(u, v)}{c_2(u, v)}\right)\right] > 0$$

In other words, the model 1 is favored over the model 2, if its log-likelihood values are significantly larger.

Vuong proposed the following statistic

$$m_i = \log\left(\frac{c_1(u,v)}{c_2(u,v)}\right) \quad , \quad i = 1,\dots,n$$

Then $\mathbf{m} = (\mathrm{m}_1,\dots,\mathrm{m}_n)^t$ is a random vector with expectation

$$E[\mathrm{m}] = \mu_0^{\mathrm{m}} = (\mu_1^{\mathrm{m}},\dots,\mu_n^{\mathrm{m}})^t,$$

if $h(\cdot)$ is the true probability mass function. If both models are equally close to the true specification, it holds $\mu_0^{\mathrm{m}} = 0$. Hence, we formulate our test problem as

$$H_0 = \mu_0^{\mathrm{m}} = 0 \quad vs \quad H_1 = \mu_0^{\mathrm{m}} \neq 0 \qquad (2)$$

Based on m, Vuong defined a test statistic v

$$\mathrm{v} = \frac{\sqrt{n}\left(\frac{1}{n}\sum_{i=1}^{n} m_i\right)}{\sqrt{\frac{1}{n}\sum_{i=1}^{n}(m_i - \overline{m})^2}} \quad , \quad \overline{m} = \frac{1}{n}\sum_{i=1}^{n} m_i$$

and has shown that under H_0

$$\mathrm{v} \xrightarrow{D} N(0,1).$$

The Clarke-test is as the Vuong-test a test for model selection based on KLIC. The difference between the Vuong and the Clarke-test is the null hypothesis. The null hypothesis of the Clarke-test is:

$$H_0 : P\left[\log\left(\frac{c_1(u,v)}{c_2(u,v)}\right) > 0\right] = p$$

If the models are equivalent then p has to be 0.5.

Application: Financial data

In this section, we study two time series of seasonal data in Iran for the period from 20.03.1990 to 20.03.2007. Further notations, we denote the variable with observations converted from the two indices as follows:

B: for the variable from Gross Domestic Production.

C: for the variable from oil income index.

We choose the Normal(N), t, Clayton(C), Gumbel(G), Frank(F), BB1 and BB7-copula to compare with the GOF tests. We denote the p-values of tests base on Kendall's transform as $P^{S_n(K_n)}$ and $P^{T_n(K_n)}$ corresponding to the Carmer von Mises (S_n) and Kolmogorov Smirnov (T_n) statistics that list in table (1). We highlighted the p-values which are greater than 0.05, because we chose the 5%-level as the critical level for our hypothesis test. Table (1) reports that T copula has the highest p-values for pair (C,B) in the tests based on Kendall's transform.

Table 1. p-values of the tests based on Kendall's transform (denoted as $P^{S_n(K_n)}$ and $P^{T_n(K_n)}$) for data pair (C, B)

$p-$ value	N	T	C	G	F	$BB1$	$BB7$
$P^{S_n(K_n)}$	0.2	0.64	0.01	0	0.01	0.19	0.17
$P^{T_n(K_n)}$	0.46	0.52	0.02	0	0	0.15	0.22

Another method to get a goodness of fit test is the Vuong or Clarke-test. If we get the following table (Table (2)) as an output of our test, the t-copula fits the data best, because the t-copula has the highest score. As a further result one can see that the Gaussian copula is the second best copula family, which is not very surprising if we remember that the t-copula converges against the Gaussian copula. The score value of 6 means that in the 6 Vuong-tests the t-copula was six times better than the other copula. The negative score of the Clayton, Frank and Gumble copula is the result of single tests, i.e. the Clayton, Frank and Gumble copula are more often beaten by the comparing copula than vise versa.

Table 2. goodness of fit table based on the Vuong-test

Test	N	T	C	G	F	$BB1$	$BB7$
Vuong	3	6	-3	-3	-4	1	0

Thereby we select T copula with no doubt.

References

1. Berg, D. (2009), Copula goodness of fit testing: an overview and power comparison. The European Journal of Finance, 15(7):675-701.
2. Genest, C. and Remillard, B. (2008), Validity of the parametric bootstrap for goodness of fit testing in semiparametric models. Ann. Inst. Henri Poincare Probab. Stat., 44(6):1096-1127.
3. Genest, C., Remillard, B. and Beaudoin, D. (2009), Goodness of fit tests for copulas: a review and a power study. Insurance: Mathematics and Economics 44, 199-213.
4. Genest, C. and Rivest, L. ,(1993), Statistical inference procedures for bivariate Archimedean copulas. J. Amer. Statist. Assoc. 88(423):10341043.
5. M. Sklar, (1959), Fonctions de repartition an dimensions et leurs marges. Publ. Inst. Stat. Univ. Paris, 8, 229-231.

The 3rd workshop on
Copula and various types of dependencies
25-26, February 2015

ON PROPERTIES OF NESTED ARCHIMEDEAN COPULAS FOR d-MONOTONE GENERATORS

REZAPOUR*, M.

[1] *Department of Statistics, Faculty of Mathematics and Computer, Shahid Bahonar University of Kerman, Kerman, Iran;*
mohsenrzp@gmail.com

Abstract. In this paper, we present some weaker conditions than those exists in the literature under which a partially nested Archimedean copula with two nesting levels is still a copula. We also obtain the density function of partially nested Archimedean copula C with two nesting levels and d_0 child copulas and state certain conditions under which a partially nested Archimedean copula with arbitrary nesting levels is indeed a copula.

1. Introduction

The Archimedean copula is a copula that is convenient in practice, flexible and appropriate for a variety of joint distributions. (See e.g. Joe (1997) and Nelsen (2006).) These copulas have many desirable properties and have a simple and closed form. They can be used to model many data sets and goodness of fit tests for this class also exist. Archimedean copula were used by Rezapour et. al. (2013a) and Rezapour and Alamatsaz (2014) to study a $(n-k+1)$-out-of-n system with dependent components. Rezapour et. al. (2013b) also considered some reliability properties of a system whose components are distributed according to an Archimedean copula. As mentioned by Marshall and Olkin (1988), this class has a close relation to Laplace Transforms

2010 *Mathematics Subject Classification.* Primary 60E05; Secondary 60K10.
Key words and phrases. Generator derivatives, Likelihood-based inference, Nested Archimedean copula.

* Speaker.

(LT). Nested Archimedean copula, as an extension of Archimedean copula, has been discussed by many authors (see Joe,1997 and McNeil,2008). McNeil (2008) gave certain conditions under which a nested Archimedean copula is indeed a copula. Here, following the work of McNeil and Nešlehová (2009), we present some weaker conditions under which a nested Archimedean copula is still a copula.

2. Main results

We recall that a *copula associated* with a multivariate distribution function(df) F is a df $C : [0,1]^d \mapsto [0,1]$ satisfying

$$F(\mathbf{x}) = C(F_1(x_1), \ldots, F_d(x_d)),$$

where, for $i = 1, ldots, d$, the F_i's are univariate marginal df's. A copula C_ψ is said to be an Archimedean copula if it can be expressed as

$$C_\psi(u_1, \ldots, u_d) = \psi\left(\sum_{i=1}^{d} \psi^{-1}(u_i)\right), \tag{2.1}$$

where $\psi : \Re_+ \mapsto [0,1]$ is a d-monotone function ($d \geq 2$), (i.e. $(-1)^k \psi^{(k)}(x) \geq 0$, $k = 0, 1, \ldots, d-2$, and $(-1)^{d-2}\psi^{(d-2)}$ is non-increasing and convex) such that $\psi(0) = 1$, and $\lim_{x\to\infty} \psi(x) = 0$. ψ is called the generator function of the copula. For more details, see [5]. If the generator of an Archimedean copula is completely monotone, i.e. $(-1)^k \psi^{(k)}(x) \geq 0$, $k = 0, 1, \ldots$, we can rewrite the Archimedean copula in (2.1) as

$$C_\psi(u_1, \ldots, u_d) = \int_0^\infty \prod_{i=1}^{d} G^\alpha(u_i) dM_\psi(\alpha), \tag{2.2}$$

where $G(x) = \exp\Big(-\psi^{-1}(x)\Big)$ and $M_\psi(.)$ is the df of a non-negative random variable with LT ψ (see [1], p. 93).

From [**?**], a partially nested Archimedean copula C with two nesting levels and d_0 child copulas (PNAC$(2, d_0, d)$), is given by

$$C(\mathbf{u}) = C_0\Big(C_1(\mathbf{u}_1), \ldots, C_{d_0}(\mathbf{u}_{d_0})\Big), \quad \mathbf{u} = (\mathbf{u}_1, \ldots, \mathbf{u}_{d_0})^T, \tag{2.3}$$

where each copula $C_j, j \in \{0, \ldots, d_0\}$, is Archimedean with generator ψ_j, that is,

$$C_j(\mathbf{u}_j) = \psi_j\Big(\psi_j^{-1}(u_{j1}) + \cdots + \psi_j^{-1}(u_{jd_j})\Big) = \psi_j\Big(t_j(\mathbf{u}_j)\Big), \tag{2.4}$$

where $t_j(\mathbf{u}_j) = \sum_{k=1}^{d_j} \psi_j^{-1}(u_{jk})$, $\psi_j : [0,\infty] \to [0,1]$ is continuous and $\dot{\psi}_{0i}(x) := \psi_0^{-1}(\psi_i(x)) \in$, $i = 1,\ldots,d_0$, satisfy the following assumptions.

Assumption 1: For $s = 1,\ldots,d_0$

(1) ψ_0 and $\psi_s(x)$ are completely monotone;
(2) $\dot{\psi}_{0i}(x) \in \hbar_\infty = \{\psi; \psi(0) = 0, \psi(\infty) = \infty, (-1)^{j-1}\psi^{(j)} \geq 0, j \geq 1\}$;
(3) ψ_0 and $\psi_s(x)$ satisfy the boundary conditions of an Archimedean copula generator.

Rezapour [10] shows that if the generators of the PNAC$(2, d_0, d)$ satisfy the following assumptions, then it is steel a copula. These conditions are weaker than those proposed in Assumption 2. Relaxing the condition on the generators of the Archimedean copula results in an increase in the number of PNAC$(2, d_0, d)$ and this, in turn, provides more distribution functions for modeling data sets.

Assumption 2: For $s = 1,\ldots,d_0$

(1) ψ_0 is d-monotone and ψ_s are d_s-monotone;
(2) $\dot{\psi}_{0s}(x) \in \hbar_{d_s}$, where

$$\hbar_k = \{\psi; \psi(0) = 0, \psi(\infty) = \infty, (-1)^{j-1}\psi^{(j)} \geq 0, 1 \leq j \leq k-2, \psi^{(k-2)} \\ \textit{is increasing and concave}\};$$

(3) ψ_0 and $\psi_s(x)$ satisfy the boundary conditions of an Archimedean copula generator.

The density function of a PNAC$(2, d_0, d)$, when ψ_0 is a m-monotone function, $m \geq d+1$ was obtained in [10] as

$$c(\mathbf{u}) = \prod_{i=1}^{d_0}\prod_{j=1}^{d_i}(\psi_i^{-1})'(u_{ij}) \sum_{\substack{k_i\in\{1,\ldots,d_i\}\\ i=1,\ldots,d_0}} \prod_{s=1}^{d_0} a_{s,d_s k_s}(t_s(\mathbf{u}_s))\psi_0^{(k^*_{d_0})}\Big(\sum_{s=1}^{d_0}\dot{\psi}_{0s}(t_s(\mathbf{u}_s))\Big), \quad (2.5)$$

where $k^*_{d_0} = k_1 + \cdots + k_{d_0}$,

$$a_{s,nk}(t_s(\mathbf{u}_s)) = \sum_{\mathbf{j}\in\mathcal{P}_{n,k}} \binom{n}{j_1,\ldots,j_{n-k+1}} \prod_{l=1}^{n-k+1}\left(\frac{\dot{\psi}_{0s}^{(l)}(t_s(\mathbf{u}_s))}{l!}\right)^{j_l},$$

$$\mathcal{P}_{n,k} = \{\mathbf{j} \in \mathbb{N}_0^{n-k+1} : \sum_{i=1}^{n-k+1} i j_i = d \textit{ and } \sum_{i=1}^{n-k+1} j_i = k\}.$$

For the sake of brevity we will solely concentrate on the case $d_0 = 2$, but extensions to the general d_0 child copulas setting are straightforward. Now suppose that $\mathbf{u} = (\mathbf{u}_1, \mathbf{u}_2)$, and $\mathbf{u}_i = (u_{i1},\ldots,u_{id_i})$, $i = 1, 2$, and consider the partially nested Archimedean copula with an m-monotone generator ψ such that ($m > d_1 + d_2 + 1$), i.e.

$$C(\mathbf{u}) = \int_{\dot{\psi}_{02}(t_1(\mathbf{u}_1))+\dot{\psi}_{02}(t_2(\mathbf{u}_2))}^{\infty} \Big(1 - \frac{\dot{\psi}_{01}(t_1(\mathbf{u}_1)) + \dot{\psi}_{02}(t_2(\mathbf{u}_2))}{t}\Big)^{n-1} dF_{R_0}(t) \,. \quad (2.6)$$

By (2.5), the corresponding density function equals

$$c(\mathbf{u}) = \prod_{i=1}^{2}\prod_{j=1}^{d_i}(\psi_i^{-1})'(u_{ij}) \sum_{k_1=1}^{d_1}\sum_{k_2=1}^{d_2}\prod_{l=1}^{2} B_{d_l,k_l}(\dot{\psi}_{0l}(t_l(\mathbf{u}_l)))$$
$$\psi_0^{(k_1+k_2)}\Big(\dot{\psi}_{01}(t_1(\mathbf{u}_1)) + \dot{\psi}_{02}(t_2(\mathbf{u}_2))\Big).$$

References

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

[2] Marshall A.W. and Olkin I. (1988), Families of multivariate distributions, *J. Amer. Statist. Assoc.* 83, 834 - 841.

[3] McNeil A. (2008), Sampling nested Archimedean copulas, *J. Statist. Comput. Simul.* 78, 567 - 581.

[4] Müller A. and Scarsini M. (2005), Archimedean copula and positive dependence, *J. Mult. Anal.* 28, 1 - 12.

[5] McNeil, A.J. and Nèslehovà, J. (2009). Multivariate Archimedean copulas, d-monotone functions and L1-norm symmetric distributions. *Ann. Stat.* 37, 3059-3097.

[6] Nelsen R.B. (2006). *An Introduction to Copulas.* Springer, New York.

[7] Rezapour M., Salehi E., and Alamatsaz M.H. (2013a), Stochastic comparison of residual and past lifetimes of $(n-k+1)$-out-of-n systems with dependent components, *Communication in Satistics Theory and Methods* 42, 2185-2199.

[8] Rezapour M., Alamatsaz M.H. and Pellerey F. (2013b), Multivariate aging with Archimedean dependence structures in high dimensions, *Communication in Satistics Theory and Methods* 42, 2056-2070.

[9] Rezapour M. and Alamatsaz M.H. (2014), Stochastic comparison of lifetimes of two $(n-k+1)$-out-of-n systems with heterogeneous dependent components, *Journal of Multivariate Analysis* 130, 240-251.

[10] Rezapour, M. (2015), On the construction of nested Archimedean copulas for d-monotone generators, Statistics and Probability Letter, 101, 21-32.