ISSN:1306-3111
e-Journal of New World Sciences Academy
2012, Volume: 7, Number: 4, Article Number: 3A0058

NWSA-PHYSICAL SCIENCES
Received: February 2012
Accepted: September 2012
Series : 3A
ISSN : 1308-7304

Çiğdem Topçu
Fahrettin Arslan
Ankara University
ctopcu@science.ankara.edu.tr
arslan@science.ankara.edu.tr
Ankara-Turkey

COPULA APPROACH FOR MODELLING DEPENDENCE STRUCTURE

ABSTRACT

This paper introduces the concept of copula as a tool to describe relationships among multivariate random variables. In this study, under the assumption of data can be modelled by one of Archimedean copulas (Gumbel, Frank, Clayton), the dependence structure between two dependent random variables with Weibull marginals is modelled by using copula.

Keywords: Copula, Archimedean Copulas, Nonparametric Estimation, Dependence, Kendall's Tau

BAĞIMLILIK YAPISINI MODELLEMEDE KAPULA YAKLAŞIMI

ÖZET

Bu çalışmada, çok boyutlu rasgele değişkenler arasındaki ilişkileri tanımlayabilmek için kullanılan kapula kavramı tanıtılmıştır. Çalışmada, verinin Archimedean kapulalardan (Gumbel, Frank, Clayton) birisi ile modellenebildiği varsayımı altında, marjinal dağılımları Weibull olan bağımlı iki rasgele değişken arasındaki bağımlılık yapısı kapula ile modellenmiştir.

Anahtar Kelimeler: Kapula, Archimedean Kapulalar, Parametrik Olmayan Tahmin, Bağımlılık, Kendall Tau

1. INTRODUCTION (GİRİŞ)

A copula is a function that links the marginals distributions to their joint distribution. Copulas first appeared in the probability metric literature. In analyzing the dependence structure using copula approaches provides many advantages such as the degree of the dependence and also the structure of the dependence. Copulas are invariant under increasing and continous transformations. For example, dependence structure with a copula does not change with logarithms of random variable. This is not true for the correlation, which is only invariant under linear transformations.

1.1. Copula Methodology (Kapula Yöntemi)

For *m* uniform random variables $U_1,U_2,...,U_m$ the corresponding the joint distribution funciton C is defined as:

$$C(u_1,u_2,...,u_m,\theta)=\Pr(U_1 \le u_1,U_2 \le u_2,...,U_m \le u_m) \qquad (1.1)$$

Here, θ is the dependence parameter.

In other words, a copula is a multivariate distribution whose marginals are all uniform over (0,1).

Combined with the fact that any continuous random variable can be transformed to be uniform over (0,1) by its probability integral transformation, copulas can be used to provide multivariate dependence structure separately from the marginal distributions[4].

A probabilistic way to define the copula is provided by the theorem of Sklar [6 and 7]. Let the continuous marginal distributions are specified by $F_1(x_1),F_2(x_2),...,F_m(x_m)$ and the joint distribution function is $H(x_1,x_2,...,x_m)$. Sklar's theorem [6] states that if H is a m-dimensional distribution function then there exists a unique m-copula C such that

$$H(x_1,x_2,...,x_m)=C(F_1(x_1),F_2(x_2),...,F_m(x_m)) \qquad (1.2)$$

The converse is also ture [6]. If C is an m-copula and $F_1(x_1),F_2(x_2),...,F_m(x_m)$ are continous then $H(x_1,x_2,...,x_m)$ is an m-dimensional distribution function with continuous marginal $F_1(x_1),F_2(x_2),...,F_m(x_m)$ such that

$$C(F_1(x_1),F_2(x_2),...,F_m(x_m))=H(x_1,x_2,...,x_m) \qquad (1.3)$$

An important property is that the copula allows the dependence structure independently from the marginal distributions. Hence, emprical marginals are usefull for estimating ther copula functions.
In this paper, we will focus on the bivariate copula function and a special class of copula is called Archimedean. Basic assumption is in this study that the data can be suitably modeled by one of Archimedean copulas.

1.2. The Archimedean Copulas and Properties (Arşimet Kapulaları ve Özellikleri)

A bivariate distribution function H(x,y) with marginals F(x) and G(y) is said to be generated by an Archimedean copula if it can be expressed in the form $C(x,y)=\phi^{-1}\left\{\phi\left\{F(x)\right\}+\phi\left\{G(y)\right\}\right\}$ for some convex, decreasing function ϕ.

ϕ denote a funciton $\Phi:[0,1]\to[0,\infty]$ which is continuous and satisfies:

- $\phi(1)=0$
- $\phi(0)=\infty$
- For all $t\in(0,1)$, $\phi'(t)<0$, we have that ϕ is decreasing
- For all $t\in(0,1)$, $\phi''(t)\geq 0$, we have that ϕ is convex.

The function ϕ has an inverse $\phi^{-1}:[0,\infty]\to[0,1]$ which has the same properties except that $\phi^{-1}(0)=1$ ve $\phi^{-1}(\infty)=0$

Table 1 gives some examples of bivariate Archimedean copulas and corresponding generator functions.

Table 1. Some examples of bivariate Archimedan copulas
(Tablo 1. İki boyutlu Archimedan kapulalar için bazı örnekler)

Family	**Generator** $\phi(t)$	**Dependence Parameter Space**	**Bivariate Copula**
Gumbel	$(-\ln t)^{\theta}$	$\theta\geq 1$	$C(u_1,u_2)=\exp[-[(-\ln u_1)^{\theta}+(-\ln u_2)^{\theta}]^{1/\theta}]$
Clayton	$t^{-\theta}-1$	$\theta>1$	$C(u_1,u_2)=((u_1)^{-\theta}+(u_2)^{-\theta}-1)^{-1/\theta}$
Frank	$-\ln\frac{e^{-\theta t}-1}{e^{-\theta}-1}$	$-\infty<\theta<\infty$	$C(u_1,u_2)=(-1/\theta)\ln\{[(1-e^{-\theta})-(1-e^{-\theta u_1})(1-e^{-\theta u_2})]/(1-e^{-\theta})\}$

The Archimedean copula presents an appealing property: each has an analytical expression that links its parameter to its related Kendall tau. The relationship between the Kendall tau and the parameter of the Archimedean copula is given in Table 2.

Table 2. Relationship between Kendall's Tau and the Parameter of Archimedean Copulas
(Tablo 2. Archimedean Kapula Parametleri ile Kendall Tau Arasındaki İlişki)

Family	Range of θ	τ
Gumbel	$\theta\in[1,\infty)$	$\frac{\theta-1}{\theta}$
Clayton	$\theta\in[0,\infty)$	$\frac{\theta}{\theta+2}$
Frank	$\theta\in(-\infty,+\infty)$	$1-\frac{4}{\theta}[1-D_1(\theta)]$

$D_1(\theta)$ is the Debye function and defined as $D_n(x)=\frac{n}{x^n}\int_0^x\frac{t^n}{e^t-1}dt,\ n>0$

2. RESEARCH SIGNIFICIANCE (ÇALIŞMANIN ÖNEMİ)

In this sudy, we focus on to illustrate the nonparametric method that has been suggested in Genest and Rivest's [2] study for estimating dependence function (also called copula function). For this purpose, we generated dependent Weibull random variables using bivariate Gumbel copula and obtain a random sample $(x_1,y_1),\ldots,(x_n,y_n)$. We assumed that this data can be suitably modeled by one of Archimedean copulas. We consider three Archimedean copulas functions: Clayton, Gumbel, Frank and use the procedure that is mentioned above to choose an appropriate copula that best fit the data. Obviously, at the end of the procedure, Gumbel copula that provide the best fit to data is expected. (Because, data are generated using Gumbel copula)

3. METHOD (METOT)

Genest and Rivest [2] suggested a nonparametric method for estimation the Archimedean copula of a pair of random variables.

The problem of specifying a probability model for independent obsarvations $(x_1,y_1),\ldots,(x_n,y_n)$ from bivariate non-Gaussian distribution funciton $H(X,Y)=T$ can be simplified by expressing H in terms of its marginals $F(x)$ and $G(y)$ and its associated dependence funciton C.

$$K(t)=\Pr[H(X,Y)\le t]=\Pr[C\{F_X(X),G_Y(Y)\}\le t] \qquad (3.1)$$

Genest and Rivest [2] provide evidence that the estimation of an Archimedean copula is uniquely determined by a function defined on the interval $(0,1)$:

$$K(t)=t-\frac{\phi(t)}{\phi'(t)} \qquad (3.2)$$

Nonparametric estimate of $K(t)$ is given by

$$K_n(t)=\sum_{j=1}^{n} I[\{T_j\le t\}]/(n+1) \qquad (3.3)$$

The estimation of K can be done in two steps. The first step consists of constructing the emprical bivariate distribution $H_n(X,Y)$. The second step consists of computing $H_n(x_i,y_i)$ for $i=1,2,\ldots,n$ and using those pseudo-observations to construct a one-dimensional emprical distribution function for K.

Table 3. Distributions of Archimedean Copulas
(Tablo 3. Archimedean Kapulaların Dağılımı)

Family	$\phi(t)$	$\phi'(t)$	$K(t)=t-\frac{\phi(t)}{\phi'(t)}$
Gumbel	$(-\ln t)^{\theta}$	$-\theta(-\ln t)^{\theta-1}\frac{1}{t}$	$t-\frac{(t\ln t)}{\theta}$
Clayton	$t^{-\theta}-1$	$-\theta t^{-\theta-1}$	$t-\frac{(t^{\theta+1}-t)}{\theta}$
Frank	$-\ln\frac{e^{-\theta t}-1}{e^{-\theta}-1}$	$\frac{\theta}{1-e^{\theta t}}$	$t-\frac{\ln\frac{e^{-\theta t}-1}{e^{-\theta}-1}}{\theta}\left(e^{\theta t}-1\right)$

3.1. Fitting a Suitable Copula to the Data (Dataya Uygun Bir Kapula Eklenmesi)

- Estimate Kendall's correlation coefficient using the nonparametric estimate of copula parameters.
- Follow the three steps below to obtain nonparametric estimate of $K(t)$, say $K_n(t)$.

i. $T_i = F_n(X_i, Y_i) = \sum_{j=1}^{n}\left[I\left\{ X_j \le X_i \,\&\, Y_j \le Y_i \right\}\right] \Big/ (n+1)$, i=1,2,…,n

ii. $K_n(t) = \dfrac{\#\{(T_i \le t)\}}{n+1}$

iii. Construct a parametric of $K(t)$ that is $K_n(t)$ using the relationship $K(t) = t - \dfrac{\phi(t)}{\phi'(t)}$

- Following Frees and Valdez [1] the selection of an Archimedean copula that fits the data beter can be done by minimizing a distance such as

$$\int \left[K_{\Phi_n}(t) - K_n(t) \right]^2 dK_n(t) \quad (3.4)$$

4. FINDINGS (BULGULAR)

We generated dependent Weibull such that $X \sim Weib(2,2)$ and $Y \sim Weib(2,2)$ using Gumbel copula with parameter $\alpha = 3.5$. Table 4 gives a summary of data. Here, as $n = 200$ is choosen arbitrarly.

Table 4. Example of data set (Tablo 4. Veri setinden örnek)

	X	Y
1	1.92175	1.92179
2	0.91639	1.00840
3	4.40737	4.35315
…		
198	1.34280	1.20311
199	1.38765	0.68966
200	2.89634	2.68194

We calculated Kendall's Tau coeficient as 0.7388 for (X,Y) as first step. Using the relationship that is shown in Table 2 parameters of copulas are calculated and results are given with Table 5.

Table 5. Nonparametric Estimation of Archimedean Copula Parameters (Tablo 5. Archimedean Kapulaların Parametrelerinin Parametrik Olmayan Tahminleri)

Family	Gumbel	Clayton	Frank
$\hat{\theta}$	3.82839	5.65679	13.43959

By following step 2 in section 3.1 , $K_n(t)$, the nonparametric estimatimates of $K(t)$, is calculated by using pseudo-observations. Obtaning the pseudo-observations is explained in step 2i. Following step 3, for parametric estimates of $K(t)$, the equation (3.2) and nonparametric estimates of parameters that are shown in Table 5 are used. Finally, selecting the suitable copula, $K_n(t)$ is compared to $K_{Gumbel}(t)$, $K_{Clayton}(t)$ and $K_{Frank}(t)$ using the square distance measure (3.4). This distance measure (3.4) is calculated 0.00200 for Gumbel copula and 0.03374, 0.00855 for Clayton and Frank copula, respectively. According to this meausure (3.4) Gumbel copula is selected. Calculations are summarized in Table 6.

Table 6. The Nonparametric and Parametric Estimates of $K_n(t)$ and Square Distance
(Tablo 6. $K_n(t)$ 'nin Parametrik ve Parametrik Olmayan Tahminleri ve Kare Uzaklık)

t	Cumulative	$K_n(t)$	$K_{Gumbel}(t)$	$K_{Clayton}(t)$	$K_{Frank}(t)$	Gumbel	Clayton	Frank
0.00001	3	0.01493	0.00004	0.00001	0.00010	0.00022	0.00022	0.00022
0.05001	20	0.09950	0.08914	0.05885	0.10097	0.00011	0.00165	0.00000
0.10001	36	0.17910	0.16016	0.11769	0.16374	0.00036	0.00377	0.00024
0.15001	49	0.24378	0.22434	0.17653	0.21923	0.00038	0.00452	0.00060
0.20001	58	0.28856	0.28409	0.23536	0.27183	0.00002	0.00283	0.00028
0.25001	67	0.33333	0.34054	0.29419	0.32311	0.00005	0.00153	0.00010
0.30001	78	0.38806	0.39436	0.35299	0.37375	0.00004	0.00123	0.00020
0.35001	90	0.44776	0.44599	0.41172	0.42407	0.00000	0.00130	0.00056
0.40001	101	0.50249	0.49575	0.47033	0.47422	0.00005	0.00103	0.00080
0.45001	110	0.54726	0.54387	0.52869	0.52428	0.00001	0.00034	0.00053
0.50001	117	0.58209	0.59054	0.58665	0.57428	0.00007	0.00002	0.00006
0.55001	132	0.65672	0.63590	0.64394	0.62422	0.00043	0.00016	0.00106
0.60001	138	0.68657	0.68007	0.70018	0.67406	0.00004	0.00019	0.00016
0.65001	146	0.72637	0.72315	0.75487	0.72374	0.00001	0.00081	0.00001
0.70001	155	0.77114	0.76522	0.80730	0.77309	0.00004	0.00131	0.00000
0.75001	163	0.81095	0.80637	0.85655	0.82183	0.00002	0.00208	0.00012
0.80001	168	0.83582	0.84664	0.90141	0.86935	0.00012	0.00430	0.00112
0.85001	177	0.88060	0.88609	0.94035	0.91450	0.00003	0.00357	0.00115
0.90001	186	0.92537	0.92478	0.97144	0.95501	0.00000	0.00212	0.00088
0.95001	194	0.96517	0.96274	0.99230	0.98641	0.00001	0.00074	0.00045
1.00000	200	1.00000	1.00000	1.00000	1.00000	0.00200	0.03374	0.00855

5. RESULT AND SUGGESTIONS (SONUÇ VE ÖNERİLER)

The statistical literature on copula modelling is stil growing. In recent years, more successful researches of copula theory and applications have been made. Most of them in finance, actuarial science and survival analysis.

There are many examples of previous studies concerning applications of copula approach. Such as Shih and Louis used Clayton copula for modelling data that has been obtained enfected by HIV patients [10]. Zheng and Klein have also used copula approach for modelling non-Hodkings lenfoma patients data [9].

In this study to illustrate the estimation procudure that has been suggested in Genest and Rivest's study [2]. The method is applicable under the assumption that dependence structure between two random variables can be modeled by Archimedean copulas. Under this assumption, we used this nonparametric method to choose the suitable parametric family of Archimedean copula for dependent random variables X and Y with Weibull marginals.

As mentioned Genest and Rivest's paper [2], the procedure is independent of marginal distributions. We considered Weibull distribution as marginals arbitrary. For other works, any continuous marginals or real data sets can be considered.

REFERENCES (KAYNAKLAR)

1. Frees W.E. and Valdez, A.E., (1997). Understanding Relationships Using Copulas. 32nd. Actuarial Research Conference, 6-8 August at University of Calgary, Albert ,Canada.
2. Genest, C. and Rivest, L.P., (1993). Statistical Inference Procedures for Bivariate Archimedean Copulas. Journal of The American Statistical Association Theory and Methods (88), No: 423.
3. Melchiori, R.M., (2003). Which Archimedean Copula is The Right One? YieldCurve.com (e-Journal)
4. Naifar, N., (2011). Modelling Dependence Structure With Archimedean Copulas And Applications iTraxx CDS index. Journal of Computational and Applied Mathematics, (235); 2459-2466.
5. Nelsen, R., (1999). An Introduction to Copulas. Springer, NewYork. Oakes, D. (1982). A Model for Association in Bivariate Survival Data. Journal of The Royal Statistical Society. Ser B. (44), 414-422.
6. Sklar, A., (1959). Fonctions de repartition a n dimensions et leurs marges. Publications de l'Institut de Statistique de lUniversite de Paris,8, 229-231.
7. Sklar, A., (1996). Random variables, distribution functions and copulas-A personal look backward and foeward- in L.Rüschendorf, B. Schweizer and M.D.Taylor (Eds.), Distributions with Fixed Marginals and Related Topics, Institute of Mathematical Statistics, Hayward, CA.
8. Zhang, D., (2005). Analysis of survival data. Lectures Notes (Modified from Dr. A. Tsiatis's lecture notes.)
9. Zheng, M. and Klein (1995). Estimates of marginal survival for dependent competing risks based on an assumed copula,Biometrika, 82 (1), 127-138.
10. Shih, J.H. and Louis, T.A., (1995). Inferences on the Association Parameter in Copula Models for Bivariate Survival Data. Biometrics 51:1384–1399.