

AKÜ FEMÜBİD 16 (2016) 021303(250-255)
DOI: 10.5578/fmbd.27971

AKU J. Sci. Eng. 16 (2016) 021303(250-255)

Araştırma Makalesi / Research Article

Comparison of Some Selection Criteria for Selecting Bivariate Archimedean Copulas

Çiğdem Topçu Gülöksüz

Bartın Üniversitesi, Fen Fakültesi, İstatistik Bölümü, 74100 Bartın.
e-mail:topcucigdem@gmail.com

Geliş Tarihi: 04.05.2016 ; Kabul Tarihi: 31.08.2016

Keywords

Copula;Dependency;Archimedean
Copula;Selection Criteria

Abstract

Commonly, while selecting an appropriate bivariate Archimedean copula function that models data, Akaike Information Criterion (AIC), Bayesian Information Criterion (BIC), and minimum distance (MD) are used as a selection criterion. In this study, the performances of these criteria for selecting copula function are investigated by some simulation studies.

İki Boyutlu Arşimedeyen Kapulalar İçin Bazı Seçim Kriterlerinin Karşılaştırılması

Anahtar Kelimeler

Kapula-Bağımlılık-Arşimedeyen
Kapulalar-Seçim Kriterleri

Özet

Genellikle, veriyi modelleyen uygun iki boyutlu Arşimedeyen kapula fonksiyonunu seçerken Akaike Bilgi Kriteri (AIC), Bayesçi Bilgi Kriteri (BIC) ile en küçük uzaklık seçim kriteri olarak kullanılır. Bu çalışmada, bu seçim kriterlerinin kapula seçimindeki performansları simülasyon çalışması ile incelenmiştir.

© Afyon Kocatepe Üniversitesi

1. Introduction

As a noun, copula means 'a link, tie, bond' and as a term, copula is a function that connects the marginal distributions to their joint distribution function. Abe Sklar first introduced copula as a term in his article in 1959. For a brief introduction to the theory, statistical properties and applications of copulas, Sklar (1973), Schweizer and Sklar (1983), Joe (1997), Frees and Valdez (1998), Nelsen (2006) can be recommended. Also, a wide range of copula families that provide various dependence structures can be found in Nelsen (2006).

Traditionally, the dependence structure between two random variables is completely described by known bivariate distributions. However, when different types of dependence structures are needed in any study, more efficient models can be used instead of standard ones. In this manner, copula functions are used to obtain such models.

A copula function allows describing dependence structure of random variables independently of their marginals. This is an important advantage that is provided by copula modeling. Copula functions also allow for asymmetric dependence unlike linear correlation coefficient. So, the

dependence structure that represents by a copula function does not change with increasing and continuous transformations. It means that copula functions are invariant under this kind of transformations (Nelsen, 2006).

All information about the dependence structure takes place in the copula. So, specifying the copula function becomes an important issue. By this purpose, estimating the copula parameter(s) is a common way for choosing a parametric copula that fits better than the other proper candidates. However, there is no certain way to choose a copula that fits data precisely. For a comprehensive research about goodness of fit tests for copulas Genest et al. (2009) can be considered. Additionally, Belgorodski (2010) is a remarkable work that provides a detailed review with applications for selecting copula functions. In this paper, the performances of the specifying ability of some selection criteria to give best possible fit to data are examined by some simulation studies.

The paper is organized as follows: in Section 2 some bivariate Archimedean copula families are introduced. The following section, estimating the copula parameter and obtaining the empirical estimate of a copula are outlined. The rest of the section, the selection criteria that are used in this work are introduced. In Section 4, some simulation studies are performed. Finally, the results are summarized and evaluated in conclusion section.

2. Some Bivariate Archimedean Copulas

According to Sklar (1959), the bivariate cumulative distribution function H of any pair (X, Y) of continuous random variables may be written in the form

$$H(x, y) = C\{F(x), G(y)\}, x, y \in \mathbb{R} \quad (1)$$

where $F(x)$ and $G(y)$ are continuous marginal distributions and C is the copula function with $C: [0,1]^2 \rightarrow [0,1]$.

Additionally, if the marginals are continuous, there is a unique copula representation (Sklar, 1959).

C is an Archimedean copula, if it can be expressed in the form of

$$C(x, y) = \phi^{-1}\{\phi\{F(x)\} + \phi\{G(y)\}\} \quad (2)$$

for some convex, decreasing function ϕ .

$\phi: [0,1] \rightarrow [0, \infty]$ is called generator function and which is continuous and has the properties below.

- i. $\phi(1) = 0$
- ii. $\phi(0) = \infty$
- iii. ϕ is decreasing, i.e. $\forall t \in (0,1), \phi'(t) < 0$
- iv. ϕ is convex, i.e. $\forall t \in (0,1), \phi''(t) \geq 0$

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

Suppose that the copula represented by (1) belongs to a family of copulas indexed by a parameter, θ , that is called the 'dependence parameter' (or copula parameter), which measures dependence between the marginals. There is a relationship between parameters of some Archimedean copulas and Kendall's tau or Spearman's rho (Nelsen, 2006). The one with Kendall's tau is preferred, here.

In this study, because of some benefits that provide such as the relationship Kendall's tau and their parameters, four Archimedean copula functions, including Gumbel (Gumbel, 1960), Clayton (Clayton, 1978), Frank (Frank, 1979) and Joe (Joe, 1997) are employed. These copula functions, corresponding parameter space and the relationship with Kendall's tau are given in Table1.

Family	Bivariate copula $C(u_1, u_2)$	Copula Parameter Space	τ
Gumbel	$\exp[-[(-\ln u_1)^\theta + (-\ln u_2)^\theta]^{1/\theta}]$	$\theta \geq 1$	$\frac{\theta - 1}{\theta}$
Clayton	$((u_1)^{-\theta} + (u_2)^{-\theta} - 1)^{-1/\theta}$	$\theta > 1$	$\frac{\theta}{\theta + 2}$
Frank	$\left(-\frac{1}{\theta}\right) \ln \left\{ \frac{(1 - e^{-\theta}) - (1 - e^{-\theta u_1})(1 - e^{-\theta u_2})}{(1 - e^{-\theta})} \right\}$	$-\infty < \theta < \infty$	$1 - \frac{4}{\theta} [1 - D_1(\theta)]$
Joe	$1 - ((u_1)^{-\theta} + (u_2)^{-\theta} - u_1^{-\theta} u_2^{-\theta})^{1/\theta}$	$\theta \geq 1$	$1 - \frac{4}{\theta} [D_j(\theta)]$

Table1. Some bivariate Archimedean copulas and the relationship with Kendall's tau

$$D_n(\theta) = \frac{n}{\theta^n} \int_0^\theta \frac{t^n}{e^{t-1}} dt, \quad n > 0, \quad D_j(\theta) = \int_0^1 \frac{(\ln(1-t^\theta)(1-t^\theta))}{t^{\theta-1}} dt, \quad n > 0$$

Gumbel copula is one of asymmetric copula functions and represents positive right tail dependence. It means that the pairs that are modeled by Gumbel copula more likely to increase together than decrease together. Clayton copula is also an asymmetric copula function and unlike Gumbel copula it represents the negative tail dependence. Frank copula is different from these two copulas. Frank is a symmetric copula function and its dependence parameter has a wide range. Joe copula is similar to the Clayton copula. Moreover, the right tail positive dependence is stronger for Joe copula.

3. Selection Procedures

The problem of selecting the copula functions among the candidates have been considered in Atkinson (1969), Atkinson(1970) and Cox (1962), Genest and Rivest (1993), Durrleman et al.(2000), Huard et al. (2006).

A random sample is drawn from the corresponding copula function C that is defined in (1). Here, the main assumption is C copula function, is a member of an Archimedean copula family.

A bivariate Archimedean copula that is defined in (2) can be uniquely determined by the function is defined on (0,1) and shown in (3) (Genest and Rivest 1993)

$$K_\phi(t) = t - \frac{\phi(t)}{\phi'(t)} \tag{3}$$

It means that a bivariate Archimedean copula function can be determined by one-dimensional $K(t)$ (Genest and Rivest 1993). In fact, $K_\phi(t)$ is the distribution function of the Archimedean copula function. Table 2 gives the studied copula functions and their distribution functions.

Table 2. Distribution functions of the Archimedean Copulas

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

This fact given in (3) is used to select suitable bivariate Archimedean copula function among the Archimedean candidates. By this purpose, the empirical estimation of $K(t)$ is compared to its theoretical estimate (Genest and Rivest 1993). Suppose that, $K_n(t)$ represents the empirical estimation of $K(t)$ and $K_\phi(t)$ represents the theoretical estimate.

The empirical estimate of $K(t)$, that is, $K_n(t)$ from random sample is obtained by using pseudo observations. Because there are not T 's, pseudo observations

$$T_i = H_n(X_i, Y_i) = \frac{\sum_{j=1}^n I\{X_j \leq X_i \& Y_j \leq Y_i\}}{n + 1}, \quad i = 1, 2, \dots, n$$

are obtained like this. Empirical estimates of $K(t)$, that is $K_n(t)$, are got by using these pseudo observations.

$$K_n(t) = \frac{\#(T_i \leq t)}{n+1}.$$

The parameters of candidate copulas are estimated by using the relationship between Kendall's tau and the studied copulas. The relationship is given in Table 1.

Using Table (2), the parametric estimate of $K(t)$, that is, $K_\phi(t)$ is constructed.

The likelihood concept is related to statistical inference. So, this concept is used in statistical literature for various purposes like fitting models to the data. Selection criteria, such as Akaike information criteria (AIC), Bayesian information criteria (BIC) are ones that serve this purpose. Also, some derivations of AIC and other information criteria can be used for model selection (Grønneberg and Hjort 2014). In this study, the selection criteria that are listed below are used to select an Archimedean copula function that gives better fit.

$AIC = -2 \sum_{i=1}^n \ln[c(u_{i,1}, u_{i,2}); \theta] + 2k$ is based on a likelihood principle and the preferred model is the lowest AIC value.

$BIC = -2 \sum_{i=1}^n \ln[c(u_{i,1}, u_{i,2}); \theta] + \ln(nk)$ and this is also based on the likelihood principle. Same as the AIC, the preferred model is with lowest BIC value.

Here, $c(x, y)$ is a density of a copula function and is defined as $c(x, y) = \frac{\partial^2 C(x, y)}{\partial x \partial y}$. And n represents sample size, k is the number of estimated parameters that model contains.

Finally, minimum distance (MD) estimation method is considered in this study to select an appropriate copula. MD estimation has been developed by Wolfowitz (Wolfowitz, 1957). It is a commonly used statistical method for fitting the model to the data. The distance between the estimated and empirical distribution is the basis of goodness of fit tests. Minimizing this distance is used as a criterion for testing the closeness. For example, Cramer-Von Mises Criterion, Kolmogorov-Smirnov Criterion and Anderson-Darling Criterion are treated as special cases of general form of distance measure. Without loss of generality, general form of this distance is preferred as a criterion.

Following Frees and Valdez (1997), minimizing the distance such as $MD = \int [K_n(t) - K_\phi(t)]^2 dK_n(t)$ is used to specify the degree of closeness of the $K_n(t)$ and $K_\phi(t)$ in this study.

4. Simulation Study

This section focused on the performances of selection criteria that were introduced in Section 3. A computer program in R (R Development Core Team 2012) is conducted for simulation studies. For studied copula functions, 100 replications of several sample sizes, $n = \{50, 100, 300\}$ were performed for three Kendall's tau values $\tau = \{0.3, 0.5, 0.7\}$ for small, medium and large dependence, respectively. For each turn, one copula function was chosen as a true copula function, and then drawn random samples of different sizes from this true copula function. Then,

Table 3. Percentages of correctly identified copula

True Copula Function	Kendall's Tau	$n_1 = 50$			$n_2 = 100$			$n_3 = 300$		
		MD	AIC	BIC	MD	AIC	BIC	MD	AIC	BIC
Gumbel	$\tau_1 = 0.3$	40	43	44	54	62	62	78	88	94
	$\tau_2 = 0.5$	60	69	69	73	84	85	95	97	100
	$\tau_3 = 0.7$	70	82	84	85	95	96	97	100	100
Clayton	$\tau_1 = 0.3$	83	83	81	92	92	93	99	100	100
	$\tau_2 = 0.5$	93	94	95	97	98	99	100	100	100
	$\tau_3 = 0.7$	95	98	98	98	99	100	100	100	100
Frank	$\tau_1 = 0.3$	38	67	68	58	81	81	89	99	100
	$\tau_2 = 0.5$	55	84	85	75	94	94	95	100	100
	$\tau_3 = 0.7$	63	94	95	80	99	99	99	100	100
Joe	$\tau_1 = 0.3$	40	69	68	54	75	75	80	90	92
	$\tau_2 = 0.5$	51	77	78	69	88	88	90	100	100
	$\tau_3 = 0.7$	56	88	88	78	95	95	99	100	100

empirical (K_n) and theoretical (K_ϕ) estimates were obtained. Finally, according to all selection criteria, the copula function that gave better possible fit was identified. Obviously, the true copula was expected to be identified. This process was replicated 100 times and reported the percentage of the true copula correctly identified by using each selection criteria. The results were summarized

5. Conclusions

In this paper, selecting an appropriate bivariate Archimedean copula function to model data by the well known selection criteria, AIC, BIC and MD were considered.

Simulations were performed to investigate the selection ability of the criteria to detect the true copula functions among the candidates.

Results showed that sample size and correlation were important features. When sample size and

dependence size increased, then the accuracy of identifying true copula also increased. However, for small sample sizes, the performances of criteria were not well.

According to the results, MD failed more times than the others while identifying the true copula functions. Particularly, at low dependence and sample size, MD had worse performances. However, AIC and BIC had same performances at any level.

Considering the randomness of these criteria, they cannot tell anything about the quality about the model. All candidates may fit poorly but even

powerful criteria. In this case, hypotheses tests may be taken account.

So, in any study you have a large sample with large dependence, the selection criteria may serve very well to specify the appropriate copula function.

REFERENCES

- Atkinson, A., 1969. A test for discriminating between models. *Biometrika* **56**, 337-341
- Atkinson, A., 1970. A method for discriminating between models. *Journal of the Royal Statistical Society, Series B32*, 323-353.
- Belgorodski, N., 2010. Selecting pair copula families for regular vines with application to the multivariate analysis of European stock market indices. Master Thesis, Technische Universität München.
- Clayton, D.G., 1978. A Model for association in bivariate life tables and its application in epidemiological studies of familial tendency in chronic disease incidence. *Biometrika*, **65**, 141–151.
- Cox, D.R., 1962. Further Results on tests of separate families of hypotheses. *Journal of the Royal Statistical Society, Series B24*, 406-424.
- Durrleman, V., Nikeghbali, A., and Roncalli, T., 2000. Which copula is the right one? Technical Report Groupe de Recherche operationnelle Credit-Lyonnois.
- Frank, M.J., 1979. On the simultaneous associativity of $F(x,y)$ and $x+y-F(x,y)$. *Aequationes Mathematicae*, **19**, 194–226.
- 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.
- 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
- Genest, C., Rémillard, B. and Beaudoin, D., 2009. goodness-of-fit tests for copulas: A review and a power study. *Insurance: Mathematics and Economics*, **44**, 199-213.
- Grønneberg, S., and Hjort N.L., 2014. The Copula Information Criteria. *Scandinavian Journal of Statistics*, **41**, 436-459.
- Gumbel, E.J., 1960. Bivariate exponential distributions. *Journal of the American Statistical Association*, **55**, 698–707.
- Huard, D., Evin, G., and Favre, A.C., 2006. Bayesian copula selection. *Computational Statistics & Data Analysis*, **51(2)**, 809-822.
- Joe, H., 1997. Multivariate Models and Dependence Concepts. Chapman & Hall Ltd.
- Nelsen, R., 2006. An Introduction to Copulas. Springer, New York.
- R-Development Core Team (2012). R: A Language and Environment for Statistical Computing (computer software). Available from: <http://www.R-project.org>.
- 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.
- Sklar, A., 1973. Random variables, joint distribution functions and copulas. *Kybernetika*, **9**, 449–460.
- Schweizer, B., and Sklar, A. 1983. *Probabilistic Metric Spaces*. North-Holland, New York.
- Wolfowitz, J., 1957. The minimum distance method. *Ann. Math. Statist.*, **28**, 75-87.