Test of symmetry for semiparametric bivariate copula

M. M. E. Abd El-Monsef* and **M. M. Seyam***

mmezzat@science.tanta.edu.eg m.seyam@science.tanta.edu.eg

* Faculty of Science, Tanta University

Abstract

The copula function is a multivariate distribution whose marginal distributions are uniformly distributed on the interval [0,1], this function called copula that ties the joint and the margins together. One important class of copula models is that of semiparametric copula models. In this paper, a semiparametric copula and its properties are introduced also a test of symmetry for semiparametric copula is derived.

Keywords: Copulas, Exchangeability, Semiparametric family, Symmetry, Measures of association

1. Introduction

The word copula was first employed in a mathematical or statistical sense by *Sklar (1959)* describing the functions that "join together" one dimensional distribution functions to form multivariate distribution functions. The theory of copulas *Joe (1997)* provides a framework for the construction of multivariate models by expressing the distribution of the data in a canonical form that models the marginal separately from the dependence structure of the data.

Copulas link joint distribution functions to their margins. Copulas are of interest to statisticians namely for two reasons: firstly, copulas are a way of studying scale-free measures of dependence, and secondly, copulas are a tool to build families of bivariate distributions with given margins *Fisher (1997)*, for further details about copulas, see *Nelsen (1999)*.

Semiparametric inference methods, based on pseudo-likelihood, have been applied to copulas by a number of authors see for example *Shih and Louis (1995)*, *Wang and Ding (2000)*, *Tsukahara (2005)* and the references there in. A semiparametric approach based on the family of bivariate Archimedean copulas *Nelsen (2006)* can be more flexible than standard parametric copula models and, at the same time, more robust to over fitting than fully non-parametric methods. Bivariate Archimedean copulas are a specific class of copulas that are uniquely determined by a unidimensional generator function.

Genest et al. (2012) proposed test for the hypothesis that the underlying copula of a continuous random pair is symmetric. The procedures are based on Cramér-von Mises and Kolmogorov-Smirnov functionals of a rank-based empirical process. *Bouzebda and Cherfi (2012)* introduced some new nonparametric statistical tests of symmetry.

In recent years, the copula theory has had a growing evolution, motivated by its applications in probability theory, statistics, finance, insurance and economics, e.g. in finance *Cherubini et al. (2004)*, risk management *McNeil et al. (2005)*, and hydrology *Salvadori et al. (2007).*

The rest of the present paper is organized as follows. In Section 2, definitions and basic properties is introduced. In Section 3, semiparametric copula is considered. In Section 4, a test of symmetry for semiparametric copula is proposed. In Section 5, measures of association are given.

2. Definition and basic properties

Consider a continuous random pair (*X*,*Y*) the joint distribution function of which is defined for all $x, y \in \mathbb{R}$ by

$$H(x,y) = P(X \leq x, Y \leq y)$$

A bivariate copula is a bivariate cumulative distribution function (cdf) with univariate uniform I margins,$C: I^2 \to I(= [0,1])$ which verifying the following properties:

a) $C(u,0) = C(0,v) = 0,\ \forall(u,v) \in I^2$
b) $C(u,1) = u \ and \ C(1,v) = v,\ \forall(u,v) \in I^2$
c) $\Delta(u_1,u_2,v_1,v_2) = C(u_2,v_2) - C(u_2,v_1) - C(u_1,v_2) + C(u_1,v_1) \geq 0,$
$\forall(u_1,u_2,v_1,v_2) \in I^4, such \ that \ u_1 \leq u_2 \ and \ v_1 \leq v_2$

The variables *X* and *Y* are said to be exchangeable if the following hypothesis holds:

$$H_0^*: \forall_{(x,y)\in\mathbb{R}^2}\ H(x,y) = H(y,x)$$

The marginal distribution functions of *X* and *Y* are defined $\forall(x,y) \in \mathbb{R}$ by

$$F(x) = P(X \leq x), \qquad G(y) = P(Y \leq y)$$

The importance of copula is described in Sklar's theorem *Sklar (1959)*: let *X* and *Y* be random variables with joint distribution function $H(x,y)$ and marginal distribution functions $F(x)$ and $G(y)$, respectively, then there exists a copula C (which is uniquely determined on Range F × Range G) such that:

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

Hence H_0^* is verified if and only if

i. $F(x) = G(x) \ for \ all \ x \in \mathbb{R}$;
ii. $C(u,v) = C(v,u) \ for \ all \ (u,v) \in [0,1]^2$.

Thus H_0^* may be rejected either because (i) or (ii) fails, or both. While Condition (i) could be validated through standard graphical or formal statistical procedures, it is not immediately clear how to test for Condition (ii).

We want to test the hypothesis

$$H_0: \forall_{(u,v)\in[0,1]^2} C(u,v) = C(v,u)$$

against the general alternative

$$H_1: \forall_{(u,v)\in[0,1]^2} C(u,v) \neq C(v,u)$$

3. Semiparametric copula

Semiparametric copula firstly introduced by *Rodríguez (1992)* and it is extensively proposed by *Amblard and Girard (2002)*. Recent literature on semiparametric copula models focused on the situation when the marginals are specified nonparametrically and the copula function is given a parametric form. Semiparametric copula is a particular case of Farlie's family introduced by *Farlie (1960)*.

One important class of copula based multivariate models is that of semiparametric multivariate copula models. Models in this class are based on parametric copulas but nonparametric marginal distributions; see for more detail *Joe (1997)* and *Nelsen (1999)*.

Let $I = [0,1]$, the semiparametric family of functions defined on I^2 by:

$$C_{\theta,\phi}(u,v) = uv + \theta\, \phi(u)\phi(v),\ \ \theta \in [-1,1] \tag{2}$$

where ϕ is a function on I.

Let (X,Y) be a random pair from the cumulative distribution function $H(x,y) = C_{\theta,\phi}(F(x),G(y))$, where $F(x)$ and $G(y)$ are respectively the cumulative distribution functions of X and Y. The estimation of the copula reduces to estimating the generating function ϕ and the parameter θ. Replacing ϕ by $\phi/\sqrt{\alpha}$ and θ by $\alpha\theta$ for any positive α leads to the same copula. When $\theta \neq 0$, introducing $s = \theta/|\theta|$ and $\psi = \sqrt{|\theta|}\phi$ yields

$$C_{\theta,\phi}(u,v) = C_{s,\psi}(u,v) = uv + s\,\psi(u)\psi(v) \tag{3}$$

where ψ satisfies the conditions of the following lemma and $s = \{-1,1\}$.

Lemma 3.1:

ϕ generates a parametric family of copulas $\{C_{\theta,\phi}(u,v),\ \theta \in [-1,1]\}$ if and only if it satisfies the following conditions:

i. $\phi(0) = \phi(1) = 0$,

ii. $\left|\phi(x) - \phi(y) \leq |x - y|\right|$ $for\ all\ (x, y) \in I^2$.

Joe (1997) introduced the definition of the Positively Quadrant Dependent (PQD), recall that X and Y are PQD, if and only if

$$\forall (x, y) \in \mathbb{R}^2, P(X \leq x, Y \leq y) \geq P(X \leq x)P(Y \leq y).$$

Amblard and Girard (2002) proposed that X and Y are PQD, if and only if

$$\theta > 0 \text{and, either } \forall u \in I,\ \phi(u) \geq 0 \text{ or } \forall u \in I, \phi(u) \leq 0.$$

Then if (X, Y) are PQD, then the copula (3) can be rewritten as:

$$C_{1,\psi}(u, v) = uv + \psi(u)\psi(v) \tag{4}$$

where ψ is a non negative function satisfying the conditions of Lemma 3.1.

4. Test of symmetry for semiparametric copula

Genest et al. (2012) introduced a test for bivariate symmetry of copulas based on Cramér-von Mises and Kolmogorov-Smirnov functional of the rank based empirical copula process by applying Jasson test (*Jasson (2005)*). *Genest et al. (2012)* provided two techniques for test of symmetry, but it does not provide more specific guidance for model choice. So, a modification for Jasson test will be introduced.

There are various concepts of bivariate symmetry for copulas. *Li and Genton (2013)* introduced three types of symmetric copula; *symmetric*, *radially symmetric and jointly symmetric*:

1) A copula *C* is symmetric if:

$$C(u, v) - C(v, u) = 0, \qquad \forall (u, v) \in [0,1]^2 \tag{5}$$

Many copulas satisfy this structure of symmetry. Moreover, two random variables are exchangeable if and only if their marginal distributions are the same and their copula is symmetric.

2) A copula *C* is radially symmetric if:

$$C(u, v) - C(1 - u, 1 - v) + 1 - u - v = 0,\ \forall (u, v) \in [0,1]^2 \tag{6}$$

There exist copulas that are symmetric but not radially symmetric and, conversely, copulas that are radially symmetric but not symmetric. In particular, exchangeability does not imply radial symmetry and vice versa.

3) A copula *C* is jointly symmetric if:

$$C(u, v) + C(u, 1 - v) - u = 0 \text{ and}$$

$$C(u, v) + C(1 - u, v) - v = 0, \forall (u, v) \in [0,1]^2$$

(7)

Jointly symmetric copulas are necessarily radially symmetric, but they could be symmetric or not, for more details on bivariate symmetry concepts see *Nelsen (1993).*

The results showed that semiparametric copula introduced in Equation (4), presented symmetric copula where:

$$uv + \psi(u)\psi(v) - vu - \psi(v)\psi(u) = 0$$

But this copula doesn't represent radially symmetric, where $\psi(u)\psi(v) - \psi(1-u)\psi(1-v)$ can't equal zero. Also doesn't represent Jointly symmetric, because neither $\psi(u)\psi(v) + \psi(u)\psi(1-v)$ nor $\psi(u)\psi(v) + \psi(1-u)\psi(v)$ equal zero.

To test whether a copula is symmetric or not, the set $[0,1]^2$can be partitioned into squares of width $1/L$ for some integer $L > 2$. By counting how many of $(\hat{U}_1, \hat{V}_1), \dots, (\hat{U}_n, \hat{V}_n)$ fall in each of these squares. If H_0 is true, the counts in cells (i,j) and (j,i) should be the same.

Consider $f: [0,1]^2 \to \mathbb{R}$ and for $k, l \in \{1, \dots, L\}$, and from property (c) in Section 2 let:

$$u_1 = \frac{k-1}{L}, u_2 = \frac{k}{L}, v_1 = \frac{l-1}{L}, v_2 = \frac{l}{L}$$

Introduce the notation:

$$p_{kl}^L(f) = f\left(\frac{k}{L}, \frac{l}{L}\right) - f\left(\frac{k-1}{L}, \frac{l}{L}\right) - f\left(\frac{k}{L}, \frac{l-1}{L}\right) + f\left(\frac{k-1}{L}, \frac{l-1}{L}\right)$$

If $p_{kl}^L(C) = p_{lk}^L(C)$ for all $k, l \in \{1, \dots, L\}$, then C is symmetric.

A local test of H_0 could be based on:

$$W_{n,(k,l)}^L = p_{kl}^L(\hat{C}_n) - p_{lk}^L(\hat{C}_n) \tag{8}$$

i.e., the difference in the proportion of counts observed in cells (k,l) and (l,k) for fixed $k, l \in \{1, \dots, L\}$, $1 \leq k < l \leq L$.

Taking into account of (4), it follows that:

$$p_{kl}^L(C_{1,\psi}) = C_{1,\psi}\left(\frac{k}{L}, \frac{l}{L}\right) - C_{1,\psi}\left(\frac{k-1}{L}, \frac{l}{L}\right) - C_{1,\psi}\left(\frac{k}{L}, \frac{l-1}{L}\right) + C_{1,\psi}\left(\frac{k-1}{L}, \frac{l-1}{L}\right)$$

$$= \frac{1}{L^2} + \left(\psi\left(\frac{k}{L}\right) - \psi\left(\frac{k-1}{L}\right)\right)\left(\psi\left(\frac{l}{L}\right) - \psi\left(\frac{l-1}{L}\right)\right)$$

Then the semiparametric estimate can be introduced:

$$p_{kl}^{L}\left(\hat{C}_{1,\psi}\right) = \frac{1}{L^2} + \left(\hat{\psi}\left(\frac{k}{L}\right) - \hat{\psi}\left(\frac{k-1}{L}\right)\right)\left(\hat{\psi}\left(\frac{l}{L}\right) - \hat{\psi}\left(\frac{l-1}{L}\right)\right)$$

and by the same steps $p_{lk}^{L}\left(\hat{C}_n\right)$ can be determined.

The proposed test statistic can be written as:

$$M_n^L = (W_n^L)^T \textstyle\sum_L^{-1} W_n^L \tag{9}$$

where

$$W_n^L = \left(W_{n,(1,2)}^L, \dots, W_{n,(L-1,L)}^L\right)^T$$

and can be calculated using (8) and Σ_L is a covariance matrix.

5. Measures of association

In this section, three invariant to strictly increasing function measures of association between the components of the random pair (X, Y) are usually considered:

5.1 Kendall's Tau

Kendall's tau measure of a pair (X, Y), distributed according to $H(x, y)$, can be defined as the difference between the probabilities of concordance and discordance of two independent pairs (X_1, Y_1) and (X_2, Y_2) described by the same joint bivariate law $H(x, y)$; that is

$$\tau_{XY} = P\{(X_1 - X_2)(Y_1 - Y_2) > 0\} - P\{(X_1 - X_2)(Y_1 - Y_2) < 0\}$$

These probabilities can be evaluated by integrating over the distribution of (X_2, Y_2). So that, in terms of copula, Kendall's tau becomes to

$$\tau = 4\int_0^1 \int_0^1 C(u, v)\, dC(u, v) - 1$$

where C is the copula associated to (X, Y).

5.2 Spearman's Rho

Let (X_1, Y_1), (X_2, Y_2) and (X_3, Y_3) be three independent random vectors with a common joint distribution function $H(x, y)$. Consider the vectors $(X_1, Y_1)$and (X_3, Y_3), then the Spearman's rho coefficient associated to a pair (X, Y), distributed according to $H(x, y)$, is defined as

$$\rho_{XY} = 3(P\{(X_1 - X_2)(Y_1 - Y_3) > 0\} - P\{(X_1 - X_2)(Y_1 - Y_3) < 0\}).$$

In terms of the copula C associated to the pair (X, Y) becomes to

$$\rho = 12 \int_0^1 \int_0^1 (C(u,v) - uv) du\, dv \tag{10}$$

Note that ρ coincides with the correlation coefficient between the uniform marginal distributions.*Amblard and Girard (2005)* proposed an expression for Spearman's rho:

$$\rho = 12\theta \left(\int_I \phi(u)\, du \right)^2 = 12 \left(\int_I \psi(u)\, du \right)^2$$

and then introduced an estimate based on the estimate $\hat{\psi}$:

$$\hat{\rho}_{SP} = 12 \left(\sum_{k \epsilon A} a_k \beta_k \right)^2$$

where $\beta_k = \int_I e_k(u) du$. The set of functions $\{e_k(w), w\epsilon I, k\epsilon A\}$ need not to be orthogonal but the condition $e_k(0) = e_k(1) = 0$ is required for all $k\epsilon A$ in order to ensure $\hat{\psi}(0) = \hat{\psi}(1) = 0$.

Amblard and Girard (2005) also proposed estimate ρ in a nonparametric copula:

$$\hat{\rho}_{NP} = \frac{6}{n(n-1)} \sum_{i=1}^{n} \sum_{j=1}^{n} 1\{u_j < u_i\,, v_j < v_i\} - \frac{3}{2}$$

where$1\{\cdot\}$ is the indicator function.

5.3 Schweizer and Wolff's σ

If we replace the function $C(u.v) - uv$ in (10) by its absolute value, then we obtain Schweizer and Wolff's σ, given by

$$\sigma = 12 \int_0^1 \int_0^1 |C(u.v) - uv| dudv$$

References

[1] Amblard, C. and Girard, S. (2002). *"Symmetry and dependence properties within a semiparametricfamily of bivariate copulas"*. Nonparametric Statistics, 14 (6), 715-727.

[2] Amblard, C. and Girard, S. (2005). *"Estimation procedures for a semiparametric family of bivariate copulas". Journal of Computational and Graphical Statistics, 14 (2), 1-15.*

[3] Bouzebda, S. and Cherfi, M. (2012).*"Test of symmetry based on copula function"*. Journal of Statistical Planning and Inference, 142, 1262-1271.

[4] Cherubini, U., Luciano, E., Vecchiato, W. (2004). *"Copula methods in finance"*. Wiley: Chichester.

[5] Genest, C. and Nešlehová, J. and Quessy, J. (2012). *"Tests of symmetry for bivariate copulas"*. Annals of the Institute of Statistical Mathematics, 64, 811-834.

[6] Farlie, D.(1960). *"The performance of some correlation coefficients for a general bivariate distribution"*. Biometrika, 47, 307–323.

[7] Fisher, N. (1997). *"Copulas"*. In: Kotz, S., Read, C.B., Banks, D.L. (Eds.), Encyclopedia of Statistical Sciences, Update Vol. 1. Wiley, New York, 159-163.

[8] Jasson, S. (2005). *"L'asymétrie de la dépendance,quel impact sur la tarification?"*. Technical report, AXA Group Risk Management, Paris, France.

[9] Joe, H. (1997). *"Multivariate models and dependence concepts"*. Monographs on Statistics and Applied Probability, 73. Chapman & Hall, London.

[10] Li, B. and Genton, M. (2013). *"Nonparametric identification of copula structures"*. Journal of the American Statistical Association, 2013.

[11] McNeil, A., Frey, R., Embrechts, P. (2005). *"Quantitative risk management: Concepts, techniques and tools"*. Princeton University Press: Princeton, NJ.

[12] Nelsen, R. (1993). *"Some concepts of bivariate symmetry"*. Journal of Nonparametric Statistics, 3, 95-101.

[13] Nelsen, R. (1999). *"An introduction to copulas"*. Springer, New York.

[14] Nelsen, R. (2006). *"An introduction to copulas (2nd ed.)"*. Springer, New York.

[15] Rodríguez-Lallena, J. (1992). *"Estudio de la compabilidad y diseno de nuevas familias en la teoria de cópulas"*. Aplicaciones. Tesis doctoral, Universidad de Granada.

[16] Salvadori, G., de Michele, C., Kottegoda, N., Rosso, R. (2007). *"Extremes in nature: An approach using copulas"*. Springer: New York.

[17] Shih, J. and Louis, T. (1995). *"Inferences on the association parameter in copula models for bivariate survival data"*. Biometrics, 51(4), 1384-1399.

[18] Sklar, A. (1959). "*Fonctions de répartitionà n dimensions et leursmarges"*. Publications del'Institutde Statistique del'Université de Paris 8, 229-231.

[19] Tsukahara, H. (2005). *"Semiparametric estimation in copula models"*. Canad. J. Statist., 33(3), 357-375.

[20] Wang, W. and Ding, A. (2000). *"On assessing the association for bivariate current status data"*. Biometrika, 87(4), 879-893.