MULTIVARIATE MEASURES OF DEPENDENCE BASED ON COPULAS

Stanisław Heilpern

Abstract. The paper is devoted to the multivariate measures of dependence. In contrast to the classical approach, where the pairs of variables are studied, we investigate the dependence of more than two variables. We mainly consider the measures based on copulas. These are the multivariable generalizations of the known coefficients of such correlation as Spearman's rho, Kendall's tau, Blomquist's beta and Gini's gamma. We present the definitions, the constructions and the basic properties of such multivariate measures of dependence. The case of large number of dimension, greater than two, presents more complications. We have several different versions of such generalization in this case and the lower bound of the values of such measures of dependence are close to zero. We also study the multivariate tail dependences. The last part of the paper is devoted to the estimation of multivariable versions of Spearman's rho coefficient.

Keywords: multivariate measures of dependence, copulas, tail dependences, estimation.

JEL Classification: C02, C13, C46.

DOI: 10.15611/me.2014.10.02.

1. Introduction

In many practical investigations, when we study the dependence between the pair of variables, we use some coefficients of correlation, e.g. Spearman's rho or Kendall's tau. In contrast of such an approach, we will investigate the dependence of more than two variables in this paper. We study the measures of dependence based mainly on copulas mainly and we generalize the known coefficients of correlation to more dimensions.

The multivariable case, when the dimension is greater than two, is significantly different from the two-dimensional case. For instance, three pairwise strict, negative dependent (countermonotonic) random variable do not exist. This fact implies that these measure of dependence do not take value

---

Stanisław Heilpern
  Department of Statistics, Wrocław University of Economics, Komandorska Street 118/120, 53-345 Wrocław, Poland.
  E-mail: stanislaw.heilpern@ue.wroc.pl

−1 and for the greater dimensions the lower bound of the values of such measures is close to zero.

First we introduce the basic notion and properties connected with copulas and multivariable measures of dependence. Next we present the generalizations of the basic coefficients of correlation: Spearman's rho, Kendall's tau, Blomquist's beta and Gini's gamma. We also study the multivariate tail dependences and the estimations of multivariate measures of dependence, mainly Spearman's tau.

2. Copulas

Let $\mathbf{X} = (X_1, \ldots, X_d)$ be a random vector, where $d \geq 2$. The dependent structure of $\mathbf{X}$ can be described by the copula. Copula C is the link between marginal F_i, where $i = 1, \ldots, d$, and the joint F cumulative distribution functions [Genest, MacKay 1986; Nelsen 2006]:

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

We assume that these cumulative distribution functions are continuous, so the copula is univocally determined in this case.

The independent joint distribution is described by the following, simple copula:

$$\Pi(\boldsymbol{u}) = u_1 \cdot \ldots \cdot u_d,$$

where $\boldsymbol{u} = (u_1, \ldots, u_d)$. For two random variables X_1, X_2 all the copulas satisfy the following inequalities:

$$W(u_1, u_2) \leq C(u_1, u_2) \leq M(u_1, u_2),$$

where copulas $W(u_1, u_2) = \max(u_1 + u_2 - 1, 0)$ and $M(u_1, u_2) = \min(u_1, u_2)$ are called respectively a lower and upper Frechet-Hoeffding bounds. If random variables X_1, X_2 have copula M then they are strict, positive dependent (comonotonic) and for copula W we obtain the strict, negative dependence (countermonotonic).

The upper bound M can be generalized to higher dimensions:

$$M(\boldsymbol{u}) = \min(u_1, \ldots, u_d).$$

We obtain the d-dimensional copula in this case. But the lower bound

$$W(\boldsymbol{u}) = \max(u_1 + \ldots + u_d - n + 1, 0)$$

is not the copula for $d > 2$. Every d-dimensional copula C satisfies the following relations:

$$W(\boldsymbol{u}) \leq C(\boldsymbol{u}) \leq M(\boldsymbol{u}).$$

There is a fundamental difference between the two cases, and the greater number of dimensions. When $d > 2$ the function $W(\mathbf{u})$ is not the copula and there do not exist three random vectors pairwise strict, negative dependent.

The copula:

$$C(u) = \Phi\left(\Phi_1^{-1}(u_1),\dots,\Phi_d^{-1}(u_d)\right),$$

where Φ is the cumulative distribution function of d-dimensional elliptical distribution and Φ_i are the marginal elliptical cumulative distribution functions is called a elliptical copula. So, we can obtain the Gaussian, t-Student or logistic copulas.

The d-dimensional copula C can be treated as a cumulative distribution function focuses on the $[0, 1]^2$ with the marginal uniform distributions, because

$$C(\boldsymbol{u}) = P(U_1 \le u_1, \dots, U_d \le u_d),$$

where $U_i = F_i(X_i)$. Its survival function is equal

$$\overline{C}(u) = P(U_1 > u_1, \dots, U_d > u_d).$$

Using the marginal survival functions $\overline{F}_i(x) = P(X_i > x)$ we can define the survival copula $\check{C}$ in the following way

$$\overline{F}(x_1,\dots,x_d) = \check{C}\left(\overline{F}_1(x_1),\dots,\overline{F}_d(x_d)\right).$$

So, we have

$$\check{C}(\boldsymbol{u}) = P(1 - U_1 \le u_1, \dots, 1 - U_d \le u_d) = \overline{C}(1 - u_1, \dots, 1 - u_d).$$

We can order the copulas using the following two relations (Schmid et al. 2010):

$$C_1 \preccurlyeq C_2 \iff C_1(\mathbf{u}) \le C_2(\mathbf{u}),$$

$$C_1 \preccurlyeq_C C_2 \iff C_1(\mathbf{u}) \le C_2(\mathbf{u}) \text{ and } \overline{C}_1(\mathbf{u}) \le \overline{C}_2(\mathbf{u}).$$

These orders are based on the natural order of copulas and its survival functions.

The Archimedean copula [Nelsen 2006]:

$$C(\mathbf{u}) = \varphi^{-1}(\varphi(u_1) + \dots + \varphi(u_d))$$

is a simple example of a copula induced by the generator φ. The generator φ is a continuous strictly decreasing function from $[0, 1]$ to $[0, \infty]$ such that

$\varphi(0) = \infty$ and $\varphi(1) = 0$ and its inverse φ^{-1} is completely monotonic, i.e. $(-1)^k \dfrac{d^k}{dt^k} \varphi^{-1}(t) \geq 0$ for all t and $k = 0, 1, 2, \ldots$

3. Basic properties of the measures of dependence

Let $\boldsymbol{\Theta}_d$ be the class of all d-dimensional random vectors. A measure of dependence p is a functional

$$p: \boldsymbol{\Theta}_d \to D \subset R.$$

Thus it assigned a real number to every random vector $\mathbf{X}$. This is a generalization of the measures of dependence, e.g. the coefficients of correlation, of two random variables to a large number of dimensions.

The Archimedean copula is generated by the one function φ only, so we usually obtain the same value of the measure of dependence for every pair of variables for many cases, i.e. $p(X_i, X_j) = p(X_k, X_l)$. We can use the two-dimensional measure of dependence p in a multivariate case, when $d > 2$:

$$p(X_1, \ldots, X_d) = p(X_1, X_2)$$

in this situation.

Another simple example of the measure of dependence is the mean p_s of two-dimensional measures of dependence p_{ij} of all pairs of variables, i.e.

$$p_s(\mathbf{X}) = \frac{2}{d(d-1)} \sum_{i<j} p_{ij}(X_i, X_j).$$

Now we present a list of properties that a "good" measure of dependence should satisfy (see [Wolf 1980; Taylor 2007; Dolati, Ubeda-Flores 2006; Schmid et al. 2010]).

W   **Well-definedness:** The measure p is well-defined for every random vector $\mathbf{X}$ and it is univocally determined by the copula, i.e. $p(\mathbf{X}) = p(C)$. So, such a measure is invariant with respect to its marginal distributions.

P   **Invariance with respect to permutations:** For every permutation π we obtain

$$p(X_1, \ldots, X_d) = p(X_{\pi(1)}, \ldots, X_{\pi(d)}).$$

N   **Normalization**

     N1   If random variables X_i are independent then $p(\mathbf{X}) = p(\Pi) = 0$.
     N2   If random variables X_i are strict, positive dependent, i.e. they have copula M, then $p(\mathbf{X}) = p(M) = 1$.

M Monotonicity

> **M1** If $C_1 \preccurlyeq C_2$ then $p(C_1) \leq p(C_2)$.
>
> **M2** If $C_1 \preccurlyeq_C C_2$ then $p(C_1) \leq p(C_2)$.

T Behavior under transformation: For every monotonic transformation I_i we have

$$p(X_1, \dots , X_d) = p(I_1(X_1), \dots , I_d(X_d)).$$

C Continuity: If $\mathbf{X}_n$ is a sequence of random vectors with copulas C_n and if $\lim\limits_{n \to \infty} C_n(\mathbf{u}) = C(\mathbf{u})$ for all $\mathbf{u}$ then $\lim\limits_{n \to \infty} p(C_n) = p(C)$.

4. The Pearson's correlation coefficient

The Pearson's correlation coefficient r is a classical, most popular two-dimensional measure of dependence. It is defined by the known formula

$$r(X_1, X_2) = \frac{\operatorname{cov}(X_1, X_2)}{\sqrt{V(X_1)V(X_2)}},$$

where covariance $\operatorname{cov}(X_1, X_2) = E(X_1 - E(X_1))(X_2 - E(X_2))$.

It does not satisfy property **W**, because it depends on the marginal distribution:

$$\operatorname{cov}(X_1, X_2) = \iint\limits_{-\infty}^{\infty} (C(F_1(x_1), F_2(x_2)) - \Pi(F_1(x_1), F_2(x_2))) dx_1 dx_2$$

and copula does not univocally determine the Pearson's coefficient. Also, it is not invariant under monotonic transformation (property **T**). For instance we have

$$r(\Phi(X_1), \Phi(X_2)) = \frac{\pi}{6} \arcsin\left(\frac{r(X_1, X_2)}{2}\right),$$

where Φ is a standard, normal distribution function.

The Pearson's coefficient does not satisfy property **N2**. Let random variables X_1 and X_2 have lognormal distribution: $X_1 \sim LN(0, 1)$, $X_2 \sim LN(0, \sigma)$, then the smallest r_{min} and greatest r_{max} values of this coefficient are equal

$$r_{\min} = r(e^Z, e^{-\sigma Z}) = \frac{e^{-\sigma} - 1}{\sqrt{(e-1)\left(e^{\sigma^2} - 1\right)}},$$

$$r_{\max} = r(e^Z, e^{\sigma Z}) = \frac{e^\sigma - 1}{\sqrt{(e-1)\left(e^{\sigma^2} - 1\right)}},$$

where $Z \sim N(0, 1)$ (see Figure 1 and [Embrechts, McNeil, Straumann 2002]).

The Pearson coefficient takes value 1 when the random variables have the same distribution, i.e. $\sigma = 1$, only and for $\sigma > 4$ all the values of this coefficient are close to 0.
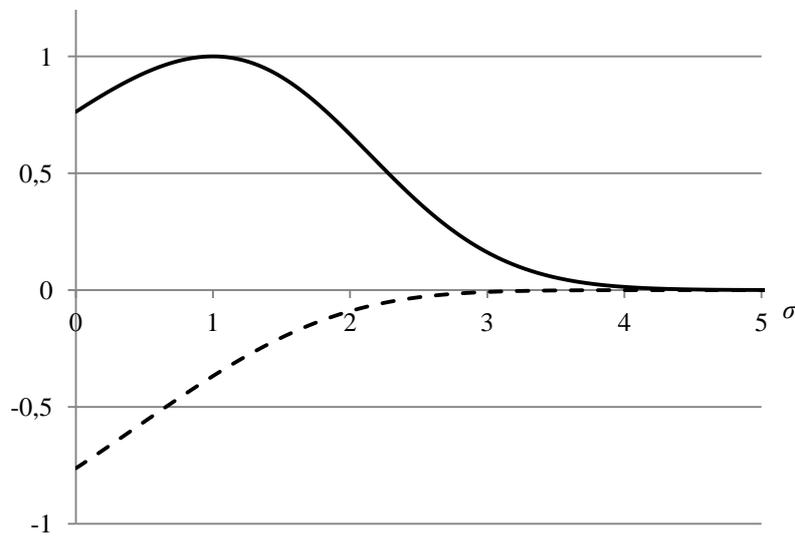

Fig. 1. The smallest and greatest values of Pearson's coefficient of correlation
for the lognormal random variables

Source: [Embrechts, McNeil, Straumann 2002].

We see that the Pearson's coefficient of correlation is not a good universal measure of dependence. It is the measure of the linear dependence only. It is not uniquely determined by a copula, so we do not present its multidimensional version.

5. Spearman's rho

Spearman's rank correlation coefficient can be defined in the two-dimensional case by the following formula:

$$\rho(X_1, X_2) = \frac{cov(F_1(X_1), F_2(X_2))}{\sqrt{V(F_1(X_1))V(F_2(X_2))}}.$$

We can present the equivalent versions of it [Nelsen 1996; Schmid et al. 2010]:

$$\rho(C) = \frac{\int_0^1 \int_0^1 C(u_1, u_2)du_1 du_2 - \int_0^1 \int_0^1 \Pi(u_1, u_2)du_1 du_2}{\int_0^1 \int_0^1 M(u_1, u_2)du_1 du_2 - \int_0^1 \int_0^1 \Pi(u_1, u_2)du_1 du_2}$$

$$= 12 \int_0^1 \int_0^1 C(u_1, u_2)du_1 du_2 - 3 \tag{1}$$

$$= 12 \int_0^1 \int_0^1 u_1 u_2 dC(u_1, u_2) - 3.$$

Spearman's rho can be interpreted as the distance from independent copula Π. It satisfies the aforementioned properties and $\rho = \frac{6}{\pi} \arcsin\left(\frac{r}{2}\right)$ for the Gaussian copula.

We present two versions of the multivariate Spearman's rho [Schmid et al. 2010]:

$$\rho_1(C) = \frac{\int_{[0,1]^d} C(\mathbf{u})d\mathbf{u} - \int_{[0,1]^d} \Pi(\mathbf{u})d\mathbf{u}}{\int_{[0,1]^d} M(\mathbf{u})d\mathbf{u} - \int_{[0,1]^d} \Pi(\mathbf{u})d\mathbf{u}} = \frac{d+1}{2^d - d - 1}\left(2^d \int_{[0,1]^d} C(\mathbf{u})d\mathbf{u} - 1\right),$$

$$\rho_2(C) = \frac{d+1}{2^d - d - 1}\left(2^d \int_{[0,1]^d} \Pi(\mathbf{u})dC(\mathbf{u}) - 1\right).$$

These are the extensions of (1). We can also use the average of the two versions:

$$\rho_3(C) = \frac{\rho_1(C) + \rho_2(C)}{2}.$$

These versions coincide when dimension $d = 2$ or copula C is radially symmetric, i.e. $C = \check{C}$. When $d = 3$, then ρ_3 is equal to the average pairwise Spearman's rho coefficient, i.e. $\rho_3 = \rho_s = (\rho_{12} + \rho_{13} + \rho_{23})/3$, where $\rho_{ij} = \rho(X_i, X_j)$. The value of multivariate coefficients ρ_i is limited from below and above by

$$\frac{2^d - (d+1)!}{d!(2^d - d - 1)} \le \rho_i(C) \le 1.$$

The graph of the lower bound of such coefficient is presented in Figure 2.

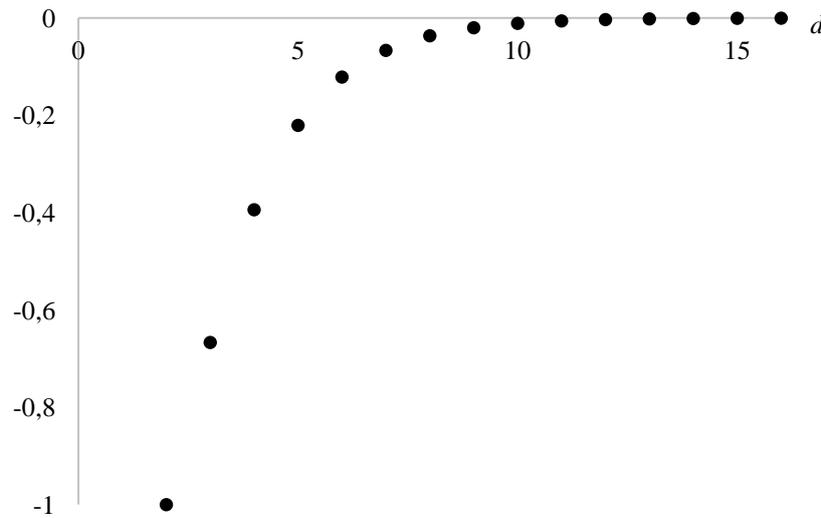Fig. 2. The graph of the lower bound of the values of the coefficients ρ_i

Source: own elaboration.

We can see that for dimension $d \ge 10$ the lower bound is close to zero. Spearman's rho takes in practice the nonnegative values only in this case. The value -1 is obtained for the bivariate case, only. This is due to the fact, that three pairwise, countermonotonic random variables do not exist.

6. Kendall's tau

Let the random vectors (X_1, X_2) and (Y_1, Y_2) be independent and identically distributed. The Kendall's tau correlation coefficient can be defined in the two-dimensional case in the following way [Nelsen 2006]:

$$\tau(X_1, X_2) = P((X_1 - Y_1)(X_2 - Y_2) > 0) - P((X_1 - Y_1)(X_2 - Y_2) < 0).$$

It is the probability of concordance minus the probability of discordance.

We can give the equivalent definition based on the copula [Nelsen 2006]:

$$\tau(C) = 4 \int\limits_{[0,1]^2} C(u,v) \, dC(u,v) - 1.$$

For bivariate Archimedean copulas, this coefficient can be calculated from the generator φ:

$$\tau(C) = 4 \int\limits_0^1 \frac{\varphi_C(t)}{\varphi_C'(t)} \, dt + 1$$

and for the Gaussian copula we have $\tau = 2 \arcsin(r) / \pi$.

Let $\mathbf{X}$, $\mathbf{Y}$ be independent and identically distributed d-dimensional random vectors. Joe proposed the following multivariable version of Kendall's tau coefficient [Joe 1990]:

$$\tau_1(\mathbf{X}) = \sum_{k=\lfloor (d+1)/2 \rfloor}^{d} w_k P\big((D_1, \ldots, D_d) \in B_{k,d-k}\big), \qquad (2)$$

where $D_i = X_i - Y_i$ and $B_{k,d-k} \subset \mathrm{R}^d$ be the subset of $\mathbf{x} \in \mathrm{R}^d$ with k negative components and $d - k$ positive. This is the family of generalizations of two-dimensional Kendall's tau. The coefficients w_k are chosen so that τ_1 satisfies "good" properties. All such choices of w_k give us the multivariate version of Kendall's tau.
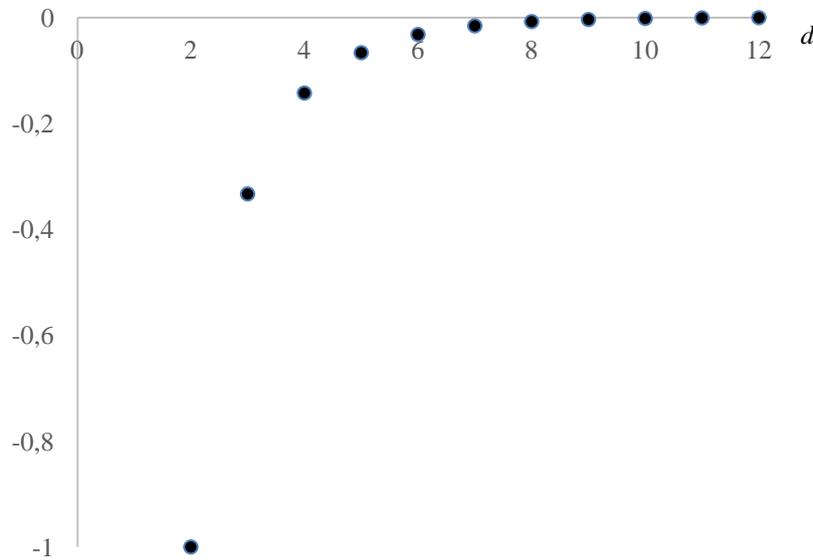

Fig. 3. The graph of the lower bound of the values of the coefficient τ_2

Source: own elaboration.

When $w_d = 1$ and $w_k = -\dfrac{1}{2^{d-1}-1}$ for $k < d$, then we have the following multivariate version of Kendall's tau coefficient [Nelsen 1996]:

$$\tau_2(C) = \frac{1}{2^{d-1}-1}\left(2^d \int\limits_{[0,1]^d} C(\boldsymbol{u})\,dC(\boldsymbol{u}) - 1 \right).$$

For $d = 3$ it is equal to the average pairwise Kendall's tau

$$\tau_s = (\tau_{12} + \tau_{13} + \tau_{23})/3,$$

where $\tau_{ij} = \tau(X_i, X_j)$ and it satisfies the following relations [Schmid et al. 2010]:

$$\frac{-1}{2^{d-1}-1} \le \tau_2(C) \le 1.$$

The graph of the lower bound of coefficient τ_2 is presented in Figure 3.

The values of the lower bound are greater than in Spearman's rho case. For $d \ge 8$ these values are practically equal to zero in this case.

7. Other measures of dependence

Blomquist's beta is a measure based on the median. Let $Z_i = X_i - \mathrm{Me}(X_i)$, the bivariate Blomquist's beta is defined by formula [Blomqvist 1950]

$$\beta(X_1, X_2) = P(Z_1 Z_2 > 0) - P(Z_1 Z_2 < 0).$$

Now we present the version of this measure based on the copula [Schmid et al. 2010]

$$\beta(C) = \frac{C\left(\frac{1}{2},\frac{1}{2}\right) - \Pi\left(\frac{1}{2},\frac{1}{2}\right) + \bar{C}\left(\frac{1}{2},\frac{1}{2}\right) - \bar{\Pi}\left(\frac{1}{2},\frac{1}{2}\right)}{M\left(\frac{1}{2},\frac{1}{2}\right) - \Pi\left(\frac{1}{2},\frac{1}{2}\right) + \bar{M}\left(\frac{1}{2},\frac{1}{2}\right) - \bar{\Pi}\left(\frac{1}{2},\frac{1}{2}\right)} = 4C\left(\frac{1}{2},\frac{1}{2}\right) - 1. \quad (3)$$

This coefficient can be interpreted as a normalized difference between copula C and the independence copula Π at $(1/2, 1/2)$.

Formula (3) let us construct the multidimensional version of Blomquist's beta coefficient [Schmid et al. 2010]:

$$\beta(C) = \frac{C\left(\frac{1}{2}\right) - \Pi\left(\frac{1}{2}\right) + \bar{C}\left(\frac{1}{2}\right) - \bar{\Pi}\left(\frac{1}{2}\right)}{M\left(\frac{1}{2}\right) - \Pi\left(\frac{1}{2}\right) + \bar{M}\left(\frac{1}{2}\right) - \bar{\Pi}\left(\frac{1}{2}\right)} = \frac{2^{d-1}}{2^{d-1} - 1}\left(C\left(\frac{1}{2}\right) + \bar{C}\left(\frac{1}{2}\right) - 2^{1-d}\right),$$

where $\frac{1}{2} = \left(\frac{1}{2}, \ldots, \frac{1}{2}\right)$.

If dimension $d = 3$, then Blomquist's beta is equal to the average pairwise values of this coefficient and it satisfies the similar relations as in Kendall's tau case [Schmid et al. 2010]:

$$\frac{-1}{2^{d-1} - 1} \leq_2 (C) \leq 1.$$

When copula C is radially symmetric, then we obtain the simple formula:

$$\beta(C) = \frac{2^d C\left(\frac{1}{2}\right) - 1}{2^{d-1} - 1}.$$

Gini's gamma coefficient is based on the extremal copulas [Nelsen 1996]:

$$\gamma(C) = 4\int_0^1\int_0^1 (M(u_1, u_2) + W(u_1, u_2))dC(u_1, u_2) - 2$$

$$= 4\left(\int_0^1\int_0^1 (A(u_1, u_2) + \bar{A}(u_1, u_2))dC(u_1, u_2) - \int_0^1\int_0^1 (A(u_1, u_2) + \bar{A}(u_1, u_2))d\Pi(u_1, u_2)\right),$$

where $A(\mathbf{u}) = (M(\mathbf{u}) + W(\mathbf{u}))/2$.

The multivariable version of Gini's gamma takes the form [Schmid et al. 2010]:

$$\gamma(C) = \frac{1}{b(d) - a(d)}\left(\int_{[0,1]^d} (A(\mathbf{u}) + \bar{A}(\mathbf{u}))dC(\mathbf{u}) - a(d)\right),$$

where

$$a(d) = \int_{[0,1]^d} (A(\mathbf{u}) + \bar{A}(\mathbf{u}))d\Pi(\mathbf{u}) = \frac{1}{d+1} + \frac{1}{2(d+1)!} + \sum_{i=0}^d (-1)^i \binom{d}{i}\frac{1}{2(i+1)!}$$

and

$$b(d) = \int\limits_{[0,1]^d} \left(A(\mathbf{u}) + \bar{A}(\mathbf{u}) \right) dM(\mathbf{u}) = 1 - \sum_{i=1}^{d-1} \frac{1}{4i!}.$$

For $d = 3$ Gini's gamma is equal to the average pairwise values of this coefficient.

8. Multivariate tail dependence

Recently, we may notice an increased interest in the analysis of extreme values, mainly in finance and insurance, e.g. insurance of the catastrophic claims [Embrechts, Kluppelberg, Mikosch 1997; Embrechts, Lindskog, McNeil 2001]. For bivariate cases, we can study dependence in the extreme values using the coefficients of tail dependence.

The **coefficient of lower tail dependence** [Sibuya 1960] is defined by the following formula:

$$\lambda_L(C) = \lim_{u \to 0} \frac{C(u,u)}{u}$$
$$= \lim_{u \to 0} P\left(X_1 \le F_1^{-1}(u) | X_2 \le F_2^{-1}(u) \right)$$
$$= \lim_{u \to 0} P\left(U_1 \le u | U_2 \le u \right),$$

where $U_i = F_i(X_i)$ and the **coefficient of upper tail dependence** by

$$\lambda_U(C) = \lim_{u \to 1} \frac{1 - 2u + C(u,u)}{1 - u} = \lim_{u \to 1} P(U_1 > u | U_2 > u).$$

Frahm [2006] proposed the following version of multivariate coefficient of lower tail dependence

$$\lambda_L(C) = \lim_{u \to 0} \frac{C(u\mathbf{1})}{1 - \bar{C}(u\mathbf{1})}$$
$$= \lim_{u \to 0} P(\max \{U_1, \ldots, U_d\} \le u | \min \{U_1, \ldots, U_d\} \le u).$$

While another version was done by Schmid and Schmidt [2007]:

$$\rho_L(C) = \lim_{p \to 0} \rho_p(C) = \lim_{p \to 0} \frac{d+1}{p^{d+1}} \int\limits_{[0,p]^d} C(\mathbf{u}) d\mathbf{u}.$$

This is based on the conditional version on Spearman's rho.

9. Estimation of the measures of dependence

In the previous section we presented the popular, theoretical version of the measures of dependence. Now we will study the empirical version of them. The empirical copula is defined by formula [Schmid et al. 2010]:

$$\hat{C}_n(\mathbf{u}) = \frac{1}{n} \sum_{j=1}^{n} \prod_{i=1}^{d} \mathbf{1}_{\{\hat{U}_{ij,n} \le u_i\}},$$

where $\hat{U}_{ij,n} = \hat{F}_{i,n}(X_{ij}) = \frac{1}{n}(\text{rank of } X_{ij} \text{ in } X_{i1}, \dots, X_{in})$ and $(\mathbf{X}_j)_{j=1,\dots,n}$ is the random sample of $\mathbf{X}$. Copula C describing the dependence structure of $\mathbf{X}$ can be estimated by the empirical copula $\hat{C}_n$. Copula $\hat{C}_n$ is treated as the empirical distribution function in this case.

The empirical version of the presented coefficients of correlation can be based on the empirical copulas. We will focus more broadly on Spearman's rho. We obtain the empirical counterpart ρ_i replacing copula C with its empirical version [Schmid, Schmidt 2007]:

$$\hat{\rho}_1 = \rho_1(\hat{C}_n) = \frac{d+1}{2^d - d - 1} \left(\frac{2^d}{n} \sum_{j=1}^{n} \prod_{i=1}^{d} \left(1 - \hat{U}_{ij,n}\right) - 1 \right)$$

$$= \frac{d+1}{2^d - d - 1} \left(\frac{2^d}{n^{d+1}} \sum_{j=1}^{n} \prod_{i=1}^{d} \left(n - R_{ij,n}\right) - 1 \right),$$

$$\hat{\rho}_2 = \rho_2(\hat{C}_n) = \frac{d+1}{2^d - d - 1} \left(\frac{2^d}{n} \sum_{j=1}^{n} \prod_{i=1}^{d} \hat{U}_{ij,n} - 1 \right)$$

$$= \frac{d+1}{2^d - d - 1} \left(\frac{2^d}{n^{d+1}} \sum_{j=1}^{n} \prod_{i=1}^{d} R_{ij,n} - 1 \right),$$

where $R_{ij,n}$ is a rank of X_{ij} in $X_{i1}, \dots, X_{in}$. Joe [1990] presented another empirical Spearman's rho:

$$\hat{\rho}_J = \frac{\dfrac{2^d}{n} \sum_{j=1}^{n} \prod_{i=1}^{d} R_{ij,n} - (n+1)^d}{\dfrac{2^d}{n} \sum_{j=1}^{n} j^m - (n+1)^d}.$$

Let us now consider a two-dimensional case, i.e. $d = 2$. Then

$$\hat{\rho}_1 = \frac{12}{n^3}\sum_{j=1}^{n}R_{1j,n}R_{2j,n} - \frac{12}{n} - 3,$$

$$\hat{\rho}_2 = \frac{12}{n^3}\sum_{j=1}^{n}R_{1j,n}R_{2j,n} - 3,$$

$$\hat{\rho}_J = \frac{12}{n^3-n}\sum_{j=1}^{n}R_{1j,n}R_{2j,n} - 3\frac{n+1}{n-1} = 1 - \frac{6\sum_{j=1}^{n}\left(R_{1j,n} - R_{2j,n}\right)^2}{n^3-n}.$$

We can see that estimators $\hat{\rho}_1$ and $\hat{\rho}_2$ are not equal in this case, because the empirical copulas are not continuous. They satisfy the inequality $\hat{\rho}_1 < \hat{\rho}_2$ and Joe's estimator $\hat{\rho}_J$ is equal to the classical empirical Spearman's rho presented for instance in Domański [1990].

Now, assume that the random variables X_1 and X_2 are comonotonic. Then the ranks are equal, i.e. $R_{1j,n} = R_{2j,n}$, so

$$\hat{\rho}_1 = \frac{n^2-6n+2}{n^2}, \quad \hat{\rho}_2 = \frac{n^2+6n+2}{n^2}$$

and $\hat{\rho}_J = 1$ in this case. The values of estimators $\hat{\rho}_1$ and $\hat{\rho}_2$ for selected values of n are presented in Table 1.

Table 1. Values of estimators $\hat{\rho}_1$ and $\hat{\rho}_2$ for comonotonic variables

n	5	10	20	50	100	500	1000	10000
$\hat{\rho}_1$	–0,1200	0,4200	0,7050	0,8808	0,9402	0,9880	0,9940	0,9994
$\hat{\rho}_2$	2,2800	1,6200	1,3050	1,1208	1,0602	1,0120	1,0060	1,0006

Source: own elaboration.

We can see that we obtain unrealistic values of these estimators for the small sample. Therefore we can use them for the large sample only, but estimator $\hat{\rho}_1$ always takes values smaller than 1 for the comonotonic varia-bles and estimator $\hat{\rho}_2$ takes greater values. Joe's estimator is a better estima-tor of Spearman's rho.

10. Conclusion

The paper is devoted to the multivariate, when dimension is greater than two, measures of dependence. The measures are based on copulas. We generalize the known coefficients of such correlation as Spearman's rho, Kendall's tau, Blomquist's beta and Gini's gamma. We can see that the case of a large number of dimensions is more complicated. We can obtain several different versions of such a generalization in this case.

The problem of the estimation of these measures of dependence is very important from the practical point of view. We showed that we could use the estimations based on the simple conversion copulas on the empirical copulas for the large sample only and that the classical estimator is better for Spearman's rho coefficient. We merely signaled the problem of estimation. The estimation issue will be discussed further in the next paper.

Acknowledgements

References

Blomqvist N. (1950). *On a measure of dependence between two random variables*. Ann. Math. Stat. 21(4). Pp. 593-600.

Dolati A., Ubeda-Flores M. (2006). *On measures of multivariate concordance*. J. Prob. Stat. Sci. 4(2). Pp. 147-164.

Domański Cz. (1990). *Testy statystyczne*. PWE. Warszawa.

Embrechts P., Kluppelberg C., Mikosch T. (1997). *Modelling External Events for Insurance and Finance*. Springer. Berlin.

Embrechts P., Lindskog F., McNeil A. (2001). *Modelling Dependence with Copulas and Applications to Risk Management*. ETH Zurich, preprint.

Embrechts P., McNeil A., Straumann D. (2002). *Correlation and dependency in risk management: properties and pitfalls*. In: M.A.H. Dempster (ed.). *Risk Management: Value at Risk and Beyond*. Cambridge University Press. Cambridge. Pp. 176-223.

Frahm G. (2006). *On the extremal dependence coefficient of multivariate distributions*. Stat. Probab. Lett. 76(14). Pp. 1470-1481.

Genest C., MacKay R.J. (1986). *The joy of copulas: Bivariate distributions with uniform marginals*. Am. Stat. 40(4). Pp. 280-285.

Joe H. (1990). *Multivariate concordance*. J. Multivar. Anal. 35(1). Pp. 12-30.

Nelsen R.B. (1996). *Nonparametric measures of multivariate association*. In: *Distribution with Fixed Marginals and Related Topics*. IMS Lecture Notes – Monograph Series 28. Institite of Mathematical Statistics. Hayward. Pp. 223-232.

Nelsen R.B. (2006). *An Introduction to Copulas* (2nd edition). Springer. New York.

Schmid F., Schmidt R. (2007). *Multivariate extensions of Spearman's rho and related statistics*. Stat. Probab. Lett. 77(4). Pp. 407-416.

Schmid F., Schmidt R., Blumentritt T., Gaißer S., Ruppert M. (2010). *Copula-Based Measures of Multivariate Association*. In: P. Jaworski, F. Durante, W. Hardle, T. Rychlik (ed.). *Copula Theory and Its Applications*. 10. Springer. Berlin.

Sibuya M. (1960). *Bivariate extreme statistics*. Ann. Inst. Math. 11(3). Pp. 195-210.

Taylor M.D. (2007). *Multivariate measures of concordance*. Ann. Inst. Stat. Math. 59(4). Pp. 789-806.

Wolff E.F. (1980). *N-dimensional measures of dependence*. Stochastica 4(3). Pp. 175-188.