

## On a measure of dependence and its application to independent component analysis

JAFAR RAHMANISHAMSI <sup>\*1</sup>, AHMAD ALIKHANI-VAFA<sup>2</sup>

<sup>1</sup>MANAGEMENT AND PLANNING ORGANIZATION, YAZD, IRAN

<sup>2</sup>DEPARTMENT OF STATISTICS, YAZD UNIVERSITY, 89175-741, YAZD, IRAN

---

Received: April 12, 2018/ Revised: July 7, 2018/ Accepted: September 3, 2018

---

**Abstract:** In this article we study a copula-based measure of dependence constructed based on the concept of average quadrant dependence. The rank-based estimator of this index and its asymptotic normality is investigated. An algorithm for independent component analysis is developed whose contrast function is the proposed dependence coefficient.

**Keywords:** Copula; Dependence measure; Independent component analysis; Test of independence. **Mathematics Subject Classification (2010):** 62F03, 62F10.

---

## 1 Introduction

Measuring the amount of dependence between random variables is an important concept in data analysis. The two popular notions of dependence are correlation, because of its simplicity, and mutual information, because of its intuitive understanding through the concept of uncertainty (Shannon and Weaver (1949)). However, correlation, because of its linear nature, and mutual information, because of its difficulty in estimation fall short in quantifying dependence. Therefore, the quest of measuring dependence in the context of an application still remains active and a lot of research devoted to the definition of measures of dependence. Schweizer and Wolff (1981) and Rényi (1959) defined fundamental properties of a measure of dependence. The necessity for dependence measures first appeared in the context of independence tests. Most of dependency measures constructed based on the distance between the joint distribution and the product of the marginal distributions, or the distance between the joint density function and the product of the marginal densities (Blum et al. (1961), Feuerverger and Mureika (1977), Hoeffding (1941), and Rosenblat (1975)). Statistical independence between random variables has become an effective tool in many signal processing

---

\*Corresponding author: jrahmanishamsi@yahoo.com

applications such as the independent component analysis (ICA), see, e.g., Bach and Jordan (2002) and Shen et al. (1949). This analysis aims at finding a transformation (usually linear) of a vector of observations, such that the transformed vector has independent components. To this end, one minimizes an objective function (also called contrast function (Comon (1994)) which measures the degree of dependence between the transformed components. In this paper, we study (In Section 2) the properties of a measure of dependence which first appeared in Behboodian et al. (2005). In Section 3, the rank-based estimator of this measure is derived and a test of independence is performed. Simulation results provided to illustrate the performance of new index in testing independence. In Section 4, an algorithm for independent component analysis is developed, whose contrast function derived from the proposed dependence measure. The proposed algorithm is compared with the commonly used ICA algorithms through some simulation results.

## 2 The proposed measure

Let  $(X, Y)$  be a pair of continuous random variables with the joint distribution function  $H(x, y) = P(X \leq x, Y \leq y)$  and the univariate marginal distributions  $F(x) = P(X \leq x)$ ,  $G(y) = P(Y \leq y)$ . The random variables  $X$  and  $Y$  are said to be positively quadrant dependent (PQD) if  $H(x, y) - F(x)G(y) \geq 0$  for all  $(x, y)$  in  $\mathbb{R}^2$  and negatively quadrant dependent (NQD) by reversing the sense of the inequality (Lehmann (1966)). So, in a sense, the expression  $H(x, y) - F(x)G(y)$  measures “local” quadrant dependence at each point  $(x, y)$  in  $\mathbb{R}^2$ . Sklar’s Theorem (Sklar (1959)) guarantees the existence of a unique function  $C$  such that, for all  $(x, y) \in \mathbb{R}^2$ ,  $H(x, y) = C\{F(x), G(y)\}$ . For a complete discussion of copulas, see Nelsen (2006). For two random variables  $X$  and  $Y$  with the associated copula  $C$ , consider a class of dependence measures of the form

$$\delta_p(C) = \lambda_B \int_0^1 \int_0^1 |C(u, v) - uv|^p dB(u, v), \quad (1)$$

where  $p \geq 1$ ,  $B$  is a fixed copula and  $\lambda_B$  is a constant depending on the copula  $B$ , such that  $\delta_p(C) = 1$  when  $C = M$  or  $C = W$ . The notions  $M$  and  $W$  denote the Fréchet-Hoeffding upper and lower bound copulas, respectively, which for any copula  $C$  satisfy:  $\max(u + v - 1, 0) = W(u, v) \leq C(u, v) \leq M(u, v) = \min(u, v)$  for every  $(u, v)$  in  $[0, 1]^2$ . Each of the random variables  $X$  and  $Y$  is almost surely (a.s.) an increasing (respectively, decreasing) function of the other if and only if their copula is  $M$  (respectively,  $W$ ). The general class (1) first appeared in Behboodian et al. (2005) and interpreted as the average quadrant dependence of two random variables  $U$  and  $V$  with the associated copula  $C$ , where the average is computed with respect to another copula  $B$ . Let  $\Pi(u, v) = uv$  denotes the copula of independent random variables. When  $B = \Pi$ , and  $p = 1$  this measure reduces to Schweizer and Wolff measure of dependence defined by

$$\sigma(C) = 12 \int_0^1 \int_0^1 |C(u, v) - uv| dudv, \quad (2)$$

(Schweizer and Wolff (1981)). When  $B = \Pi$ , and  $p = 2$ , the general class (1) reduces to the Hoeffding's measure of dependence (Hoeffding, (1941)) given by

$$\Phi(C) = 90 \int_0^1 \int_0^1 (C(u, v) - uv)^2 dudv. \quad (3)$$

A multivariate extension of this measure of dependence studied by Gaißer et al. (2010). In the general class (1), if we choose  $p = 2$  and the averaging copula  $B$  as the Fréchet-Hoeffding upper copula  $M(u, v) = \min(u, v)$ , then

$$\begin{aligned} \delta_2(C) &= 30 \int_0^1 \int_0^1 (C(u, v) - uv)^2 d \min(u, v) \\ &= 30 \int_0^1 (d_C(u) - u^2)^2 du. \end{aligned} \quad (4)$$

where  $d_C(u) = C(u, u)$  is the diagonal section of the copula  $C$ ; see, Nelsen (2006) for detail. The normalizing constant 30 obtained under the condition  $\delta_2(M) = 1$ . The measure  $\delta_2$  is comparable with the Hoeffding's measure of dependence and did not studied in literature. For two random variables  $X$  and  $Y$  with the copula  $C$ , the measure  $\delta_2(X, Y) = \delta_2(C)$ , satisfies the following properties:

- (1)  $\delta_2(X, Y)$  is well defined for every pair of continuous random variables.
- (2)  $\delta_2(X, Y) = \delta_2(X, Y)$ .
- (3)  $0 \leq \delta_2(X, Y) \leq 1$ .
- (4) If  $X$  and  $Y$  are independent then  $\delta_2(X, Y) = 0$ .
- (5)  $\delta_2(X, Y) = 1$ , when  $Y$  is an almost surely increasing transformation of  $X$ .
- (5)  $\delta_2(f(X), g(Y)) = \delta_2(X, Y)$ , if  $f$  and  $g$  are almost surely strictly monotone functions on the range of  $X$  and  $Y$ .
- (6) If  $\{(X_n, Y_n)\}$  is a sequence of continuous random vectors which converges weakly to a pair  $(X, Y)$  then  $\delta_2(X_n, Y_n) \rightarrow \delta_2(X, Y)$  as  $n \rightarrow \infty$ .

Table 1: Values of the Spearman's  $\rho$ ,  $\delta_2$  and the Hoeffding's measure  $\Phi$  for one parameter families of copulas Normal, Clayton, Frank and Gumbel.

| Spearman's | Normal   |        |            | Clayton  |        |            | Frank    |        |            | Gumbel   |        |            |
|------------|----------|--------|------------|----------|--------|------------|----------|--------|------------|----------|--------|------------|
| $\rho$     | $\theta$ | $\Phi$ | $\delta_2$ | $\theta$ | $\Phi$ | $\delta_2$ | $\theta$ | $\Phi$ | $\delta_2$ | $\theta$ | $\Phi$ | $\delta_2$ |
| 0.1        | 0.1      | 0.013  | 0.004      | 0.143    | 0.011  | 0.006      | 0.603    | 0.013  | 0.006      | 1.073    | 0.014  | 0.004      |
| 0.2        | 0.2      | 0.030  | 0.015      | 0.310    | 0.027  | 0.013      | 1.224    | 0.026  | 0.015      | 1.150    | 0.058  | 0.016      |
| 0.3        | 0.3      | 0.111  | 0.054      | 0.510    | 0.056  | 0.026      | 1.883    | 0.095  | 0.050      | 1.255    | 0.072  | 0.039      |
| 0.4        | 0.4      | 0.140  | 0.070      | 0.758    | 0.146  | 0.081      | 2.610    | 0.115  | 0.061      | 1.381    | 0.170  | 0.073      |
| 0.5        | 0.5      | 0.200  | 0.102      | 1.076    | 0.228  | 0.126      | 3.446    | 0.247  | 0.128      | 1.544    | 0.227  | 0.119      |
| 0.6        | 0.6      | 0.249  | 0.131      | 1.507    | 0.348  | 0.197      | 4.466    | 0.289  | 0.158      | 1.757    | 0.301  | 0.180      |
| 0.7        | 0.7      | 0.418  | 0.226      | 2.129    | 0.478  | 0.299      | 5.821    | 0.439  | 0.262      | 2.064    | 0.431  | 0.259      |
| 0.8        | 0.8      | 0.573  | 0.344      | 3.188    | 0.602  | 0.386      | 7.902    | 0.563  | 0.344      | 2.582    | 0.587  | 0.369      |
| 0.9        | 0.9      | 0.781  | 0.525      | 5.566    | 0.760  | 0.537      | 12.261   | 0.774  | 0.529      | 3.736    | 0.771  | 0.531      |
| 0.99       | 0.99     | 0.971  | 0.834      | 22.337   | 0.981  | 0.860      | 42.888   | 0.972  | 0.846      | 12.048   | 0.986  | 0.839      |

One disadvantage of the measure  $\delta_2$  is that a value of zero does not necessarily imply independence of random variables. This, however, also holds for the common measures of dependence such as Spearman's rho and Kendall's tau. Note that the diagonal section of an Archimedean copula  $C$  is given by  $d_C(u) = \phi^{[-1]}(2\phi(u))$ , where

Table 2: Values of the Spearman's  $\rho$ ,  $\delta_2$  and the Hoeffding's measure  $\Phi$  for one parameter families of copulas FGM and Ali-Mikhail-Haq

| <i>Spearman's</i> $\rho$ | <i>FGM</i> |        |            | <i>Ali-Mikhail-Haq</i> |        |            |
|--------------------------|------------|--------|------------|------------------------|--------|------------|
|                          | $\theta$   | $\Phi$ | $\delta_2$ | $\theta$               | $\Phi$ | $\delta_2$ |
| 0.05                     | 0.15       | 0.002  | 0.001      | 0.14                   | 0.003  | 0.002      |
| 0.1                      | 0.3        | 0.005  | 0.004      | 0.278                  | 0.009  | 0.005      |
| 0.15                     | 0.45       | 0.018  | 0.010      | 0.402                  | 0.018  | 0.012      |
| 0.2                      | 0.6        | 0.038  | 0.017      | 0.517                  | 0.038  | 0.021      |
| 0.25                     | 0.75       | 0.063  | 0.027      | 0.622                  | 0.055  | 0.030      |
| 0.3                      | 0.9        | 0.077  | 0.039      | 0.719                  | 0.089  | 0.038      |
| 0.33                     | 0.99       | 0.080  | 0.047      | 0.774                  | 0.092  | 0.049      |
| 0.35                     | -          | -      | -          | 0.808                  | 0.098  | 0.076      |
| 0.4                      | -          | -      | -          | 0.889                  | 0.143  | 0.103      |
| 0.45                     | -          | -      | -          | 0.962                  | 0.207  | 0.109      |
| 0.47                     | -          | -      | -          | 0.989                  | 0.219  | 0.114      |

$\phi : [0, 1] \rightarrow [0, \infty]$  is a continuous, convex and strictly decreasing function such that  $\phi(0) = 0$  and  $\phi^{[-1]}(u) = \phi^{-1}(u)$ , if  $0 \leq u \leq \phi(0)$  and  $\phi^{[-1]}(u) = 0$ , if  $\phi(0) \leq u \leq \infty$ , is the pseudo-inverse of  $\phi$ . As shown in Alsina et al (2006), if  $(X, Y)$  is a random vector having an Archimedean copula  $C$ , then  $X$  and  $Y$  are independent if and only if  $d_C(u) = d_\Pi(u) = u^2$ . Thus within the class of Archimedean copulas, we have that  $\delta_2(C) = 0$  if and only if  $C = \Pi$ . Figure 1 shows the plot of  $\delta_2$  and the Hoeffding's measure  $\Phi$  for the three models (1)  $Y = X^2 + c^{-1}Z$ , (2)  $Y = X + c^{-1}Z$ , where  $c \in [0, 10]$  and  $X, Z$  are two independent random variables having  $N(0,1)$  distribution, and (3) six families of copulas Clayton, Frank, Gumbel, normal, Farlie -Gumbel- Morgenstern (FGM) and Ali-Mikhail-Haq (AMH) (see, Nelsen (2006)) for different values of the parameter. Tables 1 and 2 show the values of the Spearman's  $\rho$ ,  $\delta_2$  and Hoeffding's measure  $\Phi$  for these families of copulas.

### 3 Statistical inference for $\delta_2$

An estimator for  $\delta_2$  could be written in terms of the ranks of a random sample  $(X_{1j}, X_{2j})$ ,  $j = 1, 2, \dots, n$ , from the random vector  $(X_1, X_2)$  with the copula  $C$ , and respective marginal distribution functions  $F_1$  and  $F_2$ . We assume that  $C$ , as well as  $F_1$  and  $F_2$ , are completely unknown. For  $j = 1, \dots, n$ , let  $(R_{1j}, R_{2j})$  denote the corresponding vectors of ranks, i.e.,  $R_{ij} = \sum_{l=1}^n \mathbb{I}\{X_{lj} \leq X_{ij}\}$ , for  $i = 1, 2$  and  $1 \leq j \leq n$ , where  $\mathbb{I}\{A\}$  denotes the indicator function of the set  $A$ . Let  $F_{in}$  be the re-scaled empirical counterpart of  $F_i$ ,  $i = 1, 2$ , i.e.,

$$F_{in}(x) = \frac{1}{n+1} \sum_{j=1}^n \mathbb{I}\{X_{ij} \leq x\},$$

and

$$C_n(u_1, u_2) = \frac{1}{n} \sum_{j=1}^n \mathbb{I}\{\widehat{U}_{1j} \leq u_1, \widehat{U}_{2j} \leq u_2\}, \quad (5)$$

with pseudo-observations  $\widehat{U}_{ij} = F_{in}(X_j) = R_{ij}/(n+1)$ ,  $i = 1, 2$ ,  $j = 1, 2, \dots, n$ , be the empirical copula (Fermanian (2004)). A "plug-in" rank-based estimator of  $\delta_2$  is given by

$$\delta_n = 30h(n) \int_0^1 (C_n(u, u) - u^2)^2 du, \quad (6)$$

where  $h(n)$  is a constant such that  $\delta_n = 1$  when the ranks coincide (perfect dependence), and  $\delta_n = 0$  when the ranks have a natural order (independence case), namely  $j_1, j_2, \dots, j_n$ . The following theorem provides the explicit form of  $\delta_n$ .

**Theorem 3.1.** *Let  $(X_{1j}, X_{2j})$ ,  $j = 1, 2, \dots, n$ , be a sample of size  $n$  from a vector  $(X_1, X_2)$  of continuous random variables with the copula  $C$  and let  $(R_{1j}, R_{2j})$ ,  $j = 1, \dots, n$ , be the corresponding vectors of ranks. Then*

$$\begin{aligned} \delta_n = & h(n) \left[ 16 - \frac{30}{n^2} \sum_{j=1}^n \sum_{k=1}^n \max\left(\frac{R_{1j}}{n+1}, \frac{R_{1k}}{n+1}, \frac{R_{2j}}{n+1}, \frac{R_{2k}}{n+1}\right) \right. \\ & \left. + \frac{20}{n} \sum_{j=1}^n \max\left(\frac{R_{1j}}{n+1}, \frac{R_{2j}}{n+1}\right) \right], \end{aligned}$$

where  $h(n) = n^2(n+1)/(n^3 + n^2 + 5n)$ .

**Proof:** For  $i = 1, 2$  and  $j = 1, 2, \dots, n$ , let  $\widehat{U}_{ij} = R_{ij}/(n+1)$ ,  $A_j(u) = \mathbb{I}\{\widehat{U}_{1j} \leq u\} \mathbb{I}\{\widehat{U}_{2j} \leq u\}$ . Let  $C_n$  be the associated empirical copula. Then one may write

$$\begin{aligned} (C_n(u, u) - u^2)^2 &= \left( \frac{1}{n} \sum_{j=1}^n (A_j(u) - u^2) \right)^2 \\ &= \frac{1}{n^2} \sum_{j=1}^n (A_j(u) - u^2) \sum_{k=1}^n (A_k(u) - u^2) \\ &= \frac{1}{n^2} \sum_{j=1}^n \sum_{k=1}^n A_j(u) A_k(u) - \frac{2u^2}{n} \sum_{j=1}^n A_j(u) + u^4. \end{aligned} \quad (7)$$

It is easy to see that

$$A_j(u) A_k(u) = \mathbb{I}\{\max(\widehat{U}_{1j}, \widehat{U}_{1k}, \widehat{U}_{2j}, \widehat{U}_{2k}) \leq u\}.$$

Upon integrating both sides of (7) and letting  $\widehat{U}_{ij} = \frac{R_{ij}}{n+1}$ , one gets the required result. By letting  $R_{1j} = R_{2j} = j$ , and  $R_{1k} = R_{2k} = k$ , we obtain the normalizing constant  $h(n)$ , as

$$h(n) = \frac{(n^3 + n^2 + 5n)}{n^2(n+1)}.$$

□

The asymptotic distributions of  $\delta_2$  can be deduced from the asymptotic behavior of the empirical copula process which has been discussed, e.g., in Fermanian et al. (2004) and Segers (2012). For a regular copula  $C$ , that is the partial derivatives  $\dot{C}_1(u, v) = \frac{\partial}{\partial u} C(u, v)$  and  $\dot{C}_2(u, v) = \frac{\partial}{\partial v} C(u, v)$  exist everywhere and are continuous, the empirical process  $\mathbb{C}_n = \sqrt{n}(C_n - C)$  converges weakly in  $L^\infty([0, 1]^2)$  to a centered Gaussian process  $\mathbb{G}_C$  which takes the form  $\mathbb{G}_C(u, v) = \mathbb{B}_C(u, v) - \dot{C}_1(u, v)\mathbb{B}_C(u, 1) - \dot{C}_2(u, v)\mathbb{B}_C(1, v)$ . The process  $\mathbb{B}_C$  is a Brownian bridge on  $[0, 1]^2$  with the covariance function  $\text{Cov}(\mathbb{B}_C(u_1, u_2), \mathbb{B}_C(v_1, v_2)) = C(u_1 \wedge v_1, u_2 \wedge v_2) - C(u_1, u_2)C(v_1, v_2)$ . The following result gives the asymptotic normality of  $\delta_n$ .

**Theorem 3.2.** *If  $C \neq \Pi$  is a regular copula then*

$$\sqrt{n}(\delta_n - \delta_2) \xrightarrow{d} Z_C, \quad (8)$$

where  $Z_C \sim N(0, \sigma_C^2)$  with

$$\sigma_C^2 = 4 \int_0^1 \int_0^1 (C(u, u) - u^2)(C(v, v) - v^2) E(\mathbb{G}_C(u, u)\mathbb{G}_C(v, v)) dudv, \quad (9)$$

**Proof:** Note that we can regard  $\delta_2 = \psi(C)$  as a functional acting on the space  $\mathcal{D}$  of c'adl'ag functions  $D : [0, 1]^2 \rightarrow \mathbb{R}$ , equipped with the sup norm. It is a Hadamard-differentiable map (van der Varrrt and Wellner (1996) at any copula  $C$ , tangentially to the subspace  $\mathcal{C} \subset \mathcal{D}$  of continuous maps. This taken to mean that there is a continuous linear functional  $\psi'_C : \mathcal{C} \rightarrow \mathbb{R}$  such that for every  $D \in \mathcal{C}$ ,  $\lim_{n \rightarrow \infty} \frac{\psi(C + h_n D_n) - \psi(C)}{h_n} = \psi'(D)$  where  $h_n \rightarrow 0$  and  $D_n \rightarrow D$ . Letting  $D_n = \mathbb{C}_n$  and  $h_n = n^{-1/2}$ , an application of the functional Delta Method (van der Varrrt and Wellner (1996)) to  $\mathbb{C}_n \rightarrow^w \mathbb{G}_C$  gives the required result. □

**Remark 3.3.** Note that under the assumption  $C \neq \Pi$ , the limiting random variable is non-degenerated. If  $C = \Pi$ , an application of the continuous mapping theorem yields

$$n\delta_n = 30h(n) \int_0^1 (\sqrt{n}(C_n(u, u) - u^2))^2 du \xrightarrow{d} 30 \int_0^1 (\mathbb{G}_\Pi(u, u))^2. \quad (10)$$

This result is important for testing of independence, i.e.,  $H_0 : C = \Pi$  versus  $H_1 : C \neq \Pi$ . We can reject  $H_0$  for large value of  $n\delta_n$ , for example if  $n\delta_n$  exceed from  $(1 - \alpha)$ -quantile of limiting distribution in equation (10) which can be estimated via simulation.

Since the main aim of the present work is to compare the proposed index  $\delta$  with the Hoeffding's  $\Phi$ , we include a simulation study to compare the test statistics constructed based on these measures. The simulated power comparisons presented in Figure 1, obtained with sample sizes  $n = 50$  and  $\alpha = 0.05$ . Every Monte Carlo experiment simulated 100,000 times, using well-known one-parameter Archimedean copulas Frank, Clayton, Gumbel and Ali-Mikhail-Haq and Non-Archimedean copulas normal and FGM as alternatives. For a given degree of association as measured by Spearman's rho, Figure 1 compares the power of the independence test statistics based on  $\delta_n$  and the Hoeffding's  $\Phi_n$  given by

$$\Phi_n = \frac{90}{n^2} \sum_{j=1}^n \sum_{k=1}^n \min\left(1 - \frac{R_{1j}}{n+1}, 1 - \frac{R_{1k}}{n+1}\right) \min\left(1 - \frac{R_{2j}}{n+1}, 1 - \frac{R_{2k}}{n+1}\right)$$

$$- \frac{1}{45n} \sum_{j=1}^n \left( \frac{R_{1j}}{n+1} \frac{R_{2j}}{n+1} \right)^2 - \frac{3(2n+1)+42}{5}, \quad (11)$$

as introduced in Gaißer et al. (2010). The results show a better performance for test of independence based on  $\delta_n$  than the Hoeffding's measure  $\Phi_n$ .

## 4 Application to ICA

Independent component analysis is an important problem in signal processing which consists of recovering unobserved signals from their observed mixtures. Assume that we observe  $d$  linear mixtures  $\mathbf{x} = (x_1, \dots, x_d)^T$  of  $d$  independent components  $\mathbf{s} = (s_1, \dots, s_d)^T$ , that is  $\mathbf{x} = A\mathbf{s}$ , where,  $A$  is  $d \times d$  mixing matrix. The independent components  $s_i$ 's are latent random variables with zero mean which cannot be observed directly. To solve this problem, independent component analysis (ICA) is the most popular method to extract the components via an optimization problem, in which the statistical dependency among them is minimized. The independent component analysis of a random vector involves of searching for a linear transformation that minimizes the statistical dependence between its components (Comon (1994)). More precisely, let  $X$  be a random vector on the space  $\mathbb{R}^m$ , the ICA problem would like to determine the matrix  $W \in \mathbb{R}^{m \times m}$  such that a new random vector

$$Y = WX \quad (12)$$

has its components  $Y_1, \dots, Y_m$  getting the smallest statistical dependency. Let  $\kappa$  be a measure of dependency, and  $X$  be a given  $m$ -dimensional random vector. We define the contrast function  $C_X$  from the space of  $m \times m$  real matrices to  $\mathbb{R}$ , as  $C_X(W) = \kappa(WX)$ , the dependency of the new random vector  $WX$ . The ICA is stated in a mathematical view as follows:

$$\text{Finding } W^* \in \mathbb{R}^{m \times m} \text{ such that } C_X(W^*) = \min_{W \in \mathbb{R}^{m \times m}} C_X(W). \quad (13)$$

In this section, we present an algorithm which we call MICA for ICA demixing. The algorithm uses the estimator of the index  $\delta_2$  given by (7) as a contrast function in demixing pairs of variables. First, consider the two-dimensional case, where the signal  $\mathbf{s}$  mixed with a  $2 \times 2$  matrix  $A$ . We assume that the matrix  $A$  is orthogonal. The problem is then reduced to finding a demixing rotation matrix

$$W(\theta) = \begin{pmatrix} \cos(\theta) & \sin(\theta) \\ -\sin(\theta) & \cos(\theta) \end{pmatrix}.$$

For the objective function, we use  $\delta_n$  given by (7) computed on  $2 \times N$  matrix  $\mathbf{y}(\theta) = W(\theta)\mathbf{x}$  of rotated samples. Given an angle  $\theta$ , the value of  $\delta_n(\mathbf{y}(\theta))$  can be computed by ranks of the vector  $\mathbf{y}(\theta)$ . The solution is then found by finding angle  $\theta$  minimizing  $\delta_n(\mathbf{y}(\theta))$ . We find such solution by searching over  $K$  values of  $\theta$  in the interval  $[0, \frac{\pi}{2}]$ . This algorithm is outlined in Algorithm 1. A  $d$ -dimensional linear transformation described by a  $d \times d$  orthogonal matrix  $W$  is equivalent to a composition of 2-dimensional

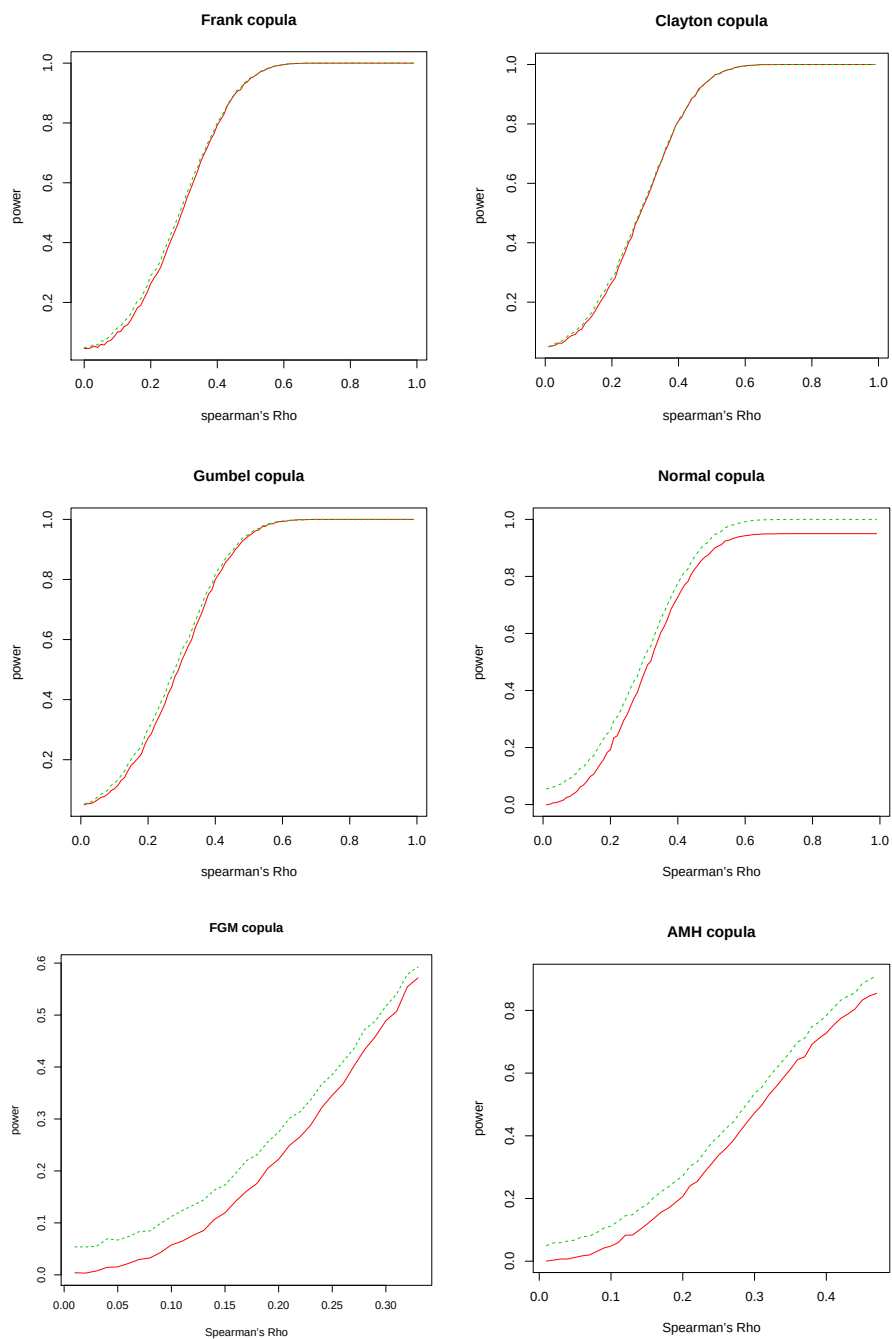


Figure 1: Power of a test of independence with the size  $\alpha = 0.05$  by using  $\delta_n$  (green line) and  $\Phi_n$  (the red line) based on a random sample of size  $n = 50$  from Frank, Clayton, Gumbel, Normal, FGM and AMH copulas



rotations; see, e.g., Comon (1994). The transformation matrix itself can be written as a product of corresponding rotation matrices,  $W = W_L \times \dots \times W_1$  where each matrix  $W_l$ ,  $l = 1, \dots, L$ , is a rotation matrix (by angle  $\theta_l$ ) for some pair of dimensions  $(i, j)$ . Thus a  $d$ -dimensional ICA problem can be solved by solving 2-dimensional ICA problems in succession. Given a current demixing matrix  $W_c = W_L \times \dots \times W_1$  and a current version of the signal  $\mathbf{x}_c = W_c \mathbf{x}$ , we find an angle  $\theta$  corresponding to Algorithm  $(\mathbf{x}_c^{(i,j)}, K)$

**Input:** A  $2 \times N$  matrix  $\mathbf{X}$  where rows are mixed signals (centered and whitened),  $k$  equispaced evaluation angles in the  $[0, \pi/2)$  interval for each of  $K$  angles  $\theta$  in the interval  $[0, \pi/2)$ ,  $\theta = \frac{\pi k}{2K}$ ,  $k = 0, \dots, K-1$ .

**Procedure:**

[1] Compute rotation matrix

$$W(\theta) = \begin{pmatrix} \cos(\theta) & \sin(\theta) \\ -\sin(\theta) & \cos(\theta) \end{pmatrix}.$$

[2] Compute rotated signals  $\mathbf{Y}(\theta) = W(\theta)\mathbf{X}$

[3] Compute  $\delta_n(\mathbf{Y}(\theta))$  sample estimate of M index.

[4] Find best angle  $\theta_m = \operatorname{argmin}_{\theta} \delta_n(\mathbf{Y}(\theta))$ .

**Output:** Rotation matrix  $\mathbf{W} = \mathbf{W}(\theta_m)$ , demixed signal  $\mathbf{X} = \mathbf{Y}(\theta_m)$ , and estimated dependence measure  $\delta_n(\mathbf{Y}(\theta))$

#### Algorithm 1: MICA algorithm

In the following we present simulation results to compare the proposed rank-based method with the commonly used ICA algorithms FastICA (Hyvärinen et al. (1999); Hyvärinen and Köster (2006)), KernelICA (Bach and Jordan (2002); Gretton et al. (2005)), RADICAL (Learned-Miller (2003)) and JADE (Cardoso and Souloumiac (1993); Cardoso (1999)). In the simulation study 9 different one-dimensional densities including: (1)  $t(3)$ , the student  $t$  distribution with 3 degree of freedom, (2)  $t(5)$ , (3) Uniform on the interval  $(0,1)$ , (4) Exponential, (5) Chi-Square, (6) Lognormal, (7) F, (8) unimodal mixture of normal and (9) multimodal mixture of normal densities are used to generating original independent sources. The shapes of the densities presented in Figure 2. The procedure for generating data is as the following: (1)  $N$  samples of each of the  $d$  sources were generated according to their density functions and placed into an  $d \times N$  matrix  $\mathbf{X}$ , (2) a random mixing matrix  $A$  was chosen, (3) a matrix  $\mathbf{Y}$  of dimension  $d \times N$  was formed as the mixture  $\mathbf{Y} = A\mathbf{X}$ , (4) the data were whitened by multiplying  $\mathbf{Y}$  by the inverse of the square root of the sample covariance matrix, yielding an  $d \times N$  matrix of whitened data  $\mathbf{Y}$ . This matrix is the input of the ICA algorithms. Each of the ICA algorithms outputs a demixing matrix  $W$  which can be applied to the matrix  $\mathbf{Y}$  to recover estimates of the independent components. To evaluate the performance of the algorithm, we compare Amari error (Amari (1996)) or blind performance index of MICA algorithm with those of the other algorithms. Let

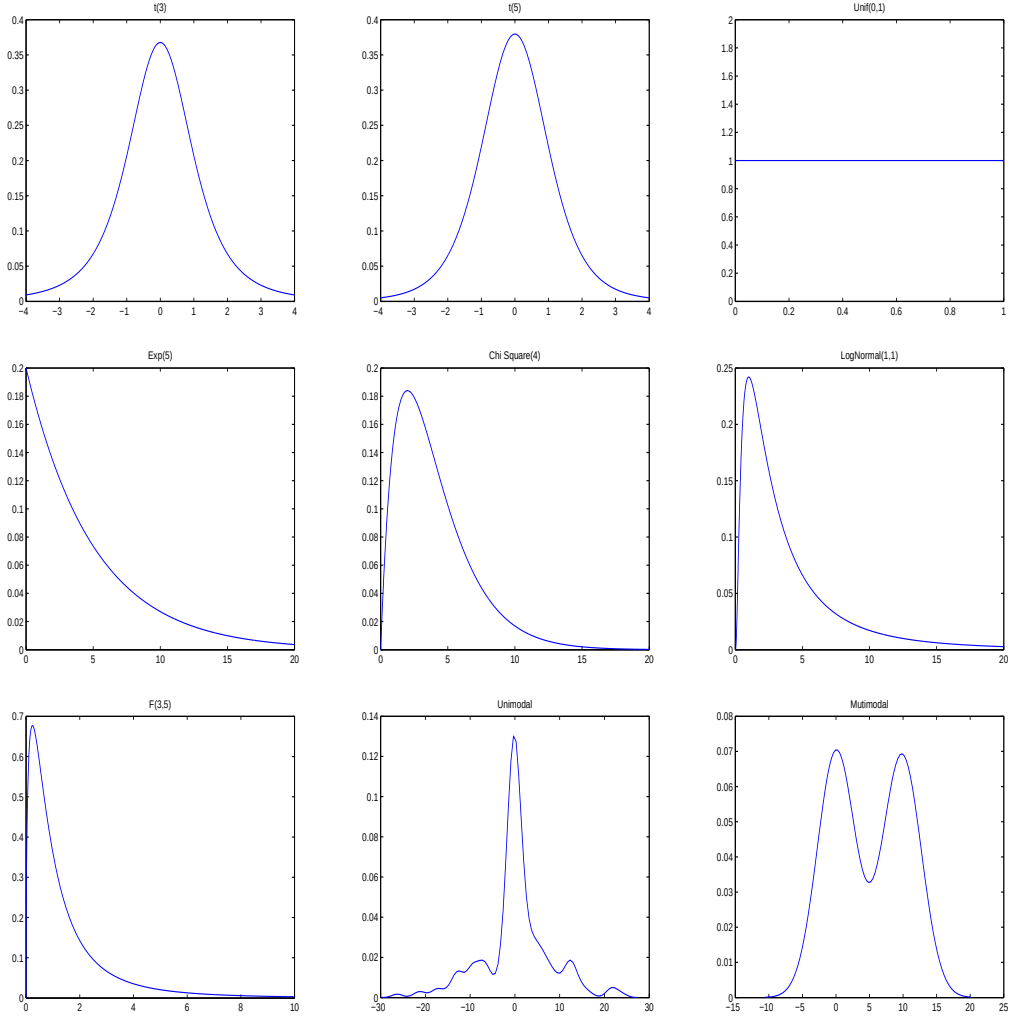


Figure 2: The shapes of the density functions used in the simulation study.

$b_{ij}$  be the entries of the performance matrix  $\mathbf{B} = WA$ . The Amari error  $r(\mathbf{B})$  measures how different matrix  $\mathbf{B}$  is from a permutation matrix, and defined by

$$r(\mathbf{B}) = \frac{1}{2d(d-1)} \left( \sum_{i=1}^d \left( \frac{\sum_{j=1}^d |b_{ij}|}{\max_j |b_{ij}|} - 1 \right) + \sum_{j=1}^d \left( \frac{\sum_{i=1}^d |b_{ij}|}{\max_i |b_{ij}|} - 1 \right) \right). \quad (14)$$

It takes value in  $[0,1]$ , with the minimum value 0 if and only if  $\mathbf{B}$  is a permutation matrix. Tables 4, 5 and 6 summarize the medians of the Amari errors for 2-dimensional, 4-dimensional and 8-dimensional cases where both sources had the same distribution. Samples from these sources were then transformed by a random rotation, and then demixed using competing ICA algorithms. As we see from Table 3, in 2-dimensional

case, when the initial sources come from the near-Gaussian-tailed distribution ( $t(3)$ ,  $t(5)$ ) and the uniform distribution, the MICA algorithm is not better than usual algorithms. But for the sources with the heavytail distributions, rank-based algorithm MICA can recover unobserved signals from their observed mixtures closer than usual algorithms. Similar results have been seen in 4-dimentional and 8-dimentional cases in the Table 4 and Table 5, respectively.

Table 3: The Amari errors (multiplied by 100) for 2-component ( $d=2$ ) ICA with 1000 samples. Each entry is the median of 100 replicates for each pdf, (1) to (10). The lowest (best) entry in each row is boldfaced.

| Num | Distribution                 | Kernel ICA | JADE         | Fast ICA    | RADICAL | MICA        |
|-----|------------------------------|------------|--------------|-------------|---------|-------------|
| 1   | $t(3)$                       | 9.43       | 7.02         | <b>5.93</b> | 7.92    | 14.94       |
| 2   | $t(5)$                       | 14.10      | <b>11.15</b> | 11.78       | 13.25   | 22.17       |
| 3   | Uniform                      | 9.84       | 62.05        | <b>8.17</b> | 8.98    | 18.43       |
| 4   | Exp                          | 8.55       | 26.55        | 12.56       | 9.02    | <b>5.24</b> |
| 5   | Chi square                   | 9.42       | 38.69        | 9.65        | 10.05   | <b>6.99</b> |
| 6   | Lognormal                    | 8.71       | 22.17        | 8.51        | 7.92    | <b>5.24</b> |
| 7   | F                            | 8.50       | 17.75        | 9.48        | 8.98    | <b>7.87</b> |
| 8   | Unimodal mixture of normal   | 8.59       | 9.58         | 9.21        | 8.98    | <b>5.09</b> |
| 9   | Multimodal mixture of normal | 9.26       | 50.63        | 9.73        | 8.98    | <b>6.11</b> |

Table 4: The Amari errors (multiplied by 100) for 4-component ( $d=4$ ) ICA with 1000 samples. Each entry is the median of 100 replicates for each pdf, (1) to (10). The lowest (best) entry in each row is boldfaced

| Num | Distribution                 | Kernel ICA | JADE         | Fast ICA | RADICAL     | MICA        |
|-----|------------------------------|------------|--------------|----------|-------------|-------------|
| 1   | $t(3)$                       | 10.11      | 8.78         | 9.49     | <b>8.03</b> | 13.53       |
| 2   | $t(5)$                       | 19.17      | <b>11.44</b> | 13.75    | 13.12       | 21.43       |
| 3   | Uniform                      | 7.20       | 35.80        | 6.89     | <b>5.51</b> | 19.35       |
| 4   | Exp                          | 5.18       | 16.28        | 10.87    | 5.15        | <b>4.17</b> |
| 5   | Chi square                   | 5.36       | 16.39        | 10.98    | 5.64        | <b>5.16</b> |
| 6   | Lognormal                    | 5.18       | 16.28        | 10.87    | 5.15        | <b>3.88</b> |
| 7   | F                            | 4.48       | 12.12        | 7.04     | 5.12        | <b>3.70</b> |
| 8   | Unimoda mixture of normal    | 5.51       | 9.37         | 11.12    | 5.20        | <b>4.67</b> |
| 9   | Multimodal mixture of normal | 4.93       | 30.75        | 5.32     | 4.99        | <b>4.50</b> |

Table 5: The Amari errors (multiplied by 100) for 8-component ( $d=8$ ) ICA with 1000 samples. Each entry is the median of 100 replicates for each pdf, (1) to (10). The lowest (best) entry in each row is boldfaced.

| Num | Distribution                 | Kernel ICA | JADE         | Fast ICA | RADICAL     | MICA        |
|-----|------------------------------|------------|--------------|----------|-------------|-------------|
| 1   | $t(3)$                       | 12.10      | 8.72         | 9.48     | <b>7.74</b> | 14.39       |
| 2   | $t(5)$                       | 20.57      | <b>10.90</b> | 12.80    | 12.00       | 22.45       |
| 3   | Uniform                      | 5.87       | 25.46        | 6.64     | <b>5.32</b> | 20.08       |
| 4   | Exp                          | 5.86       | 11.30        | 13.46    | 4.75        | <b>4.70</b> |
| 5   | Chi square                   | 5.82       | 10.90        | 12.29    | 4.41        | <b>4.35</b> |
| 6   | Lognormal                    | 4.22       | 9.33         | 7.66     | 4.12        | <b>3.98</b> |
| 7   | F                            | 3.87       | 7.93         | 5.81     | 3.86        | <b>3.21</b> |
| 8   | Unimodal mixture of normal   | 38.26      | 15.27        | 17.13    | 4.21        | <b>4.12</b> |
| 9   | Multimodal mixture of normal | 42.54      | 16.43        | 4.22     | 4.03        | <b>3.97</b> |

## References

- Alsina, C., Schweizer, B. and Frank, M.J. (2006). Associative functions: triangular norms and copulas. Singapore: World Scientific.
- Amari, S., Cichocki, A. and Yang, H.H. (1996). A new learning algorithm for blind signal separation. *Advances in Neural Information Processing Systems*, **8**, 757–763. MIT Press.
- Bach, F.R. and Jordan, M.I. (2002). Kernel independent component analysis. *The Journal of Machine Learning Research*, **3**, 1–48.
- Behboodian, J., Dolati, A. and Úbeda-Flores, M. (2005). Measures of association based on average quadrant dependence. *Journal of Probability and Statistical Science*, **3**, 161–173.
- Blum, J.R., Kiefer, J. and Rosenblatt, M. (1961). Distribution free tests of independence based on the sample distribution function. *The Annals of Mathematical Statistics*, **32**, 485–498.
- Cardoso, J.F. and Souloumiac, A. (1993). Blind beam forming for non-Gaussian signals. *In IEE Proceedings F- Radar and Signal Processing*, (Vol. 140, No. **6**, pp. 362–370).
- Cardoso, J.F. (1999). High-order contrasts for independent component analysis. *Neural Computation*, **11**(1), 157–192.
- Comon, P. (1994). Independent component analysis, a new concept?. *Signal Processing*, **36**(3), 287–314.
- Fermanian, J.D., Radulovic, D. and Wegkamp, M. (2004). Weak convergence of empirical copula processes. *Bernoulli*, **10**(5), 847–860.
- Feuerverger, A. and Mureika, R.A. (1977). The empirical characteristic function and its applications. *The Annals of Statistics*, **5**, 88–97.
- Gaißer, S., Ruppert, M. and Schmid, F. (2010). A multivariate version of Hoeffding’s phi-square. *Journal of Multivariate Analysis*, **101**(10), 2571–2586.
- Gretton, A., Bousquet, O., Smola, A. and Scholkopf, B. (2005). Measuring statistical dependence with Hilbert-Schmidt norms. *In Algorithmic Learning Theory* (pp. 63–77). Springer Berlin Heidelberg.
- Hoeffding, W. (1941). Scale-invariant correlation theory. *In The Collected Works of Wassily Hoeffding* (pp. 57–107). Springer New York.
- Hyvärinen, A. (1999). Fast and robust fixed-point algorithms for independent component analysis. *IEEE Transactions on Neural Networks*, **10**, 626–634.
- Hyvärinen, A. and Koster, U. (2006). FastISA: A fast fixed-point algorithm for independent subspace analysis. *In Proc. of ESANN*, Evere, Belgium.

- Learned-Miller, E.G. (2003). ICA using spacings estimates of entropy. *The Journal of Machine Learning Research*, **4**, 1271-1295.
- Lehmann, E.L. (1966). Some concepts of dependence. *The Annals of Mathematical Statistics*, **35**, 1137-1153.
- Nelsen, R.B. (2006). An introduction to copulas. Springer Science & Business Media.
- Rényi, A. (1959). On measures of dependence. *Acta Mathematica Hungarica*, **10**, 441-451.
- Rosenblatt, M. (1975). A quadratic measure of deviation of two-dimensional density estimates and a test of independence. *The Annals of Statistics*, **3**, 1-14.
- Segers, J. (2012). Asymptotics of empirical copula processes under non-restrictive smoothness assumptions, *Bernoulli*, **18**, 764–782.
- Shannon, C.E. and Weaver, W. (1949). The Mathematical Theory of Communication (Champaign, IL. Urbana: University of Illinois Press).
- Shannon, C.E. and Weaver, W. (1949). The Mathematical Theory of Communication. Urbana.
- Shen, H., Jegelka, S. and Gretton, A. (2009). Fast kernel-based independent component analysis. *IEEE Transactions on Signal Processing*, **57**(9), 3498–3511.
- Schweizer, B. and Wolff, E.F. (1981). On nonparametric measures of dependence for random variables. *The Annals of Statistics*, **9**, 879–885.
- Sklar, A. (1959). Fonctions de répartition à  $n$  dimensions et leurs marges. Institut de statistique de l'Université de Paris, **8**, 229–231.
- Van der Vaart, A.W. and Wellner, J.A. (1996). Weak Convergence and Empirical Processes, *Springer Verlag*.