View metadata, citation and similar papers at core.ac.uk
brought to you by CORE
provided by Repositorio da Producao Cientifica e Intelectual da Unicamp

UNIVERSIDADE ESTADUAL DE CAMPINAS
SISTEMA DE BIBLIOTECAS DA UNICAMP
REPOSITÓRIO DA PRODUÇÃO CIENTIFICA E INTELECTUAL DA UNICAMP

**Versão do arquivo anexado / Version of attached file:**

Versão do Editor / Published Version

**Mais informações no site da editora / Further information on publisher's website:**
https://www.sciencedirect.com/science/article/pii/S0047259X12002667

**DOI: 10.1016/j.jmva.2012.11.007**

**Direitos autorais / Publisher's copyright statement:**
©2013 by Elsevier. All rights reserved.

DIRETORIA DE TRATAMENTO DA INFORMAÇÃO

Cidade Universitária Zeferino Vaz Barão Geraldo
CEP 13083-970 – Campinas SP
Fone: (19) 3521-6493
http://www.repositorio.unicamp.br

Journal of Multivariate Analysis 115 (2013) 481–495

Contents lists available at SciVerse ScienceDirect

Journal of Multivariate Analysis

journal homepage: www.elsevier.com/locate/jmva

# A new index to measure positive dependence in trivariate distributions

Jesús E. García [a], V.A. González-López [a,*], R.B. Nelsen [b]

[a] Department of Statistics, University of Campinas, Rua Sérgio Buarque de Holanda, 651, Campinas, São Paulo, CEP 13083-859, Brazil

[b] Department of Mathematical Sciences, Lewis and Clark College, 0615 SW Palatine Hill Road, Portland, OR 97219, USA

ARTICLE INFO

*Article history:*
Received 20 November 2011
Available online 29 November 2012

*AMS 2000 subject classifications:*
62H20
62G30

*Keywords:*
Measure of association
Directional dependence
Copula

ABSTRACT

We introduce a new index to detect dependence in trivariate distributions. The index is based on the maximization of the coefficients of directional dependence over the set of directions. We show how to calculate the index using the three pairwise Spearman's rho coefficients and the three common 3-dimensional versions of Spearman's rho. We obtain the asymptotic distributions of the empirical processes related to the estimators of the coefficients of directional dependence and also we derive the asymptotic distribution of our index. We display examples where the index identifies dependence undetected by the aforementioned 3-dimensional versions of Spearman's rho. The value of the new index and the direction in which the maximal dependence occurs are easily computed and we illustrate with a simulation study and a real data set.

© 2012 Elsevier Inc. All rights reserved.

## 1. Introduction

In this paper we define and study an index to detect positive dependence in trivariate distributions, undetected by the existing 3-dimensional versions of Spearman's rho. The 3-dimensional versions of Spearman's rho are frequently used to develop independence tests, and to do that it is necessary to investigate the empirical copula process and the survival copula process, in order to obtain the asymptotic law of continuous functionals of the latter empirical processes, as showed in Quessy [21].

In several situations as pointed out in Gaißer and Schmid [12] the assumption of equality between the pairwise correlations allows us to use particular statistical models. In Gaißer and Schmid [12] four nonparametric tests for testing the hypothesis of equal Spearman's rho coefficients in a multivariate random vector have been proposed and the asymptotic distribution of the tests has been established as a consequence of the asymptotic behavior of the empirical copula process. To test constant correlations, for example, if we want to test if correlations of asset returns change in time, it is necessary to choose some correlation coefficient and in general the limiting distribution of the test statistic is obtained under the condition of finite fourth moments. But Wied et al. [29] presents a fluctuation test for constant correlation based on Spearman's rho that does not require any moments, where the limit distribution of the test statistic is the supremum of the absolute value of a Brownian bridge that provides critical values without any bootstrap techniques. The empirical copula process is important not only for statistics based on Spearman's rho, but also for others such as a multivariate version of Hoeffding's Phi-Square, as illustrated in Gaißer et al. [11], in which is proposed a multivariate version for Hoeffding's bivariate measure of association, Phi-Square. In addition, a nonparametric estimator is proposed and its asymptotic behavior established, based on the weak convergence of the empirical copula process.

* Corresponding author.
*E-mail addresses:* jg@ime.unicamp.br (J.E. García), veronica@ime.unicamp.br (V.A. González-López), nelsen@lclark.edu (R.B. Nelsen).

0047-259X/$ – see front matter © 2012 Elsevier Inc. All rights reserved.
doi:10.1016/j.jmva.2012.11.007

Our target is to give the foundation for the construction of a new index of trivariate dependence, which is capable of detecting dependence undetected by the traditional trivariate extensions of Spearman's rho. We present the asymptotic distribution of the empirical process related to the estimator of that index, under relatively weak conditions, as discussed in Segers [25]. We also obtain the asymptotic distributions of the empirical processes related to the estimators of the coefficients of directional dependence postulated by Nelsen and Úbeda-Flores [20]. In the sequel, several tests of independence can be formulated, using the asymptotic laws here developed and following the ideas of Genest et al. [15] and Quessy [21] or in a more specific context, would be possible to apply the ideas of Rifo and González-López [16]. The family of Spearman's rho coefficients is especially appropriate to test constant correlations for example, as was showed in Wied et al. [29]. In practice, the use of Spearman's rho correlation in that kind of tests allows us to analyze several types of data (non-elliptical data for instance) taking advantage of the robustness which is a natural property of rank-based measures.

In Section 2 we review the definitions of three well-known 3-dimensional versions of Spearman's rho and we discuss the coefficients of directional dependence introduced by Nelsen and Úbeda-Flores [20], since our index, denoted by $\rho_3^{\max}$, is based on the maximization of those coefficients over all directions. In Section 3 we introduce formally the index $\rho_3^{\max}$, we prove the main result of our paper showing that the new index can be easily written as a function of the pairwise Spearman correlations and the 3-dimensional versions of Spearman's rho. We exhibit situations in which the index $\rho_3^{\max}$ detects dependence undetected by the most common 3-dimensional versions of Spearman's rho. Theoretical properties of the index are presented in the same section. In Section 4 we show how to estimate our index using well-known estimators. In addition, in Section 5, we establish the asymptotic normality for the estimators of the coefficients of directional dependence, for the estimator of the 3-dimensional versions of Spearman's rho and for the estimator of the index $\rho_3^{\max}$. In Section 6 we compute $\rho_3^{\max}$ in different situations, a simulation study and an application to real data set. In Section 7 we emphasize the simplicity of the new index, its good properties and we stress situations in which the new index has an outstanding performance.

## 2. Preliminaries

Given a pair $(X_1, X_2)$ of continuous random variables with associated 2-copula $C$, the population version of Spearman's rho, denoted by $\rho_{12}(C)$ is defined by

$$\rho_{12}(C) = 12 \int_{I^2} C(u, v)dudv - 3, \tag{1}$$

where $I = [0, 1]$.

We omit the argument $C$ to simplify the notation when the underlying copula is understood. In the trivariate case, where $(X_1, X_2, X_3)$ is a vector of continuous random variables with 3-copula $C$, there are several generalizations of Spearman's rho. They are, (a) the average of the three pairwise measures $\rho_{12}$, $\rho_{13}$ and $\rho_{23}$, where each pairwise measure is given by Eq. (1)

$$\rho_3^*(C) = \frac{\rho_{12} + \rho_{13} + \rho_{23}}{3}, \tag{2}$$

(b) the trivariate generalizations given by Joe [17] and Nelsen [19]

$$\rho_3^-(C) = 8 \int_{I^3} C(u, v, w)dudvdw - 1, \tag{3}$$

$$\rho_3^+(C) = 8 \int_{I^3} \overline{C}(u, v, w)dudvdw - 1, \tag{4}$$

where $\overline{C}$ denotes the survival function associated with $C$, and

(c) the coefficients of directional dependence $\rho_3^{(\alpha_1,\alpha_2,\alpha_3)}(C)$ introduced by Nelsen and Úbeda-Flores [20], where $\alpha_i \in \{-1, 1\}$, given by

$$\rho_3^{(\alpha_1,\alpha_2,\alpha_3)}(C) = 8 \int_{I^3} Q_{\alpha_1,\alpha_2,\alpha_3}(u, v, w)dudvdw, \tag{5}$$

where $Q_{\alpha_1,\alpha_2,\alpha_3}(u, v, w)$ is $P(\alpha_1 X_1 > \alpha_1 u, \alpha_2 X_2 > \alpha_2 v, \alpha_3 X_3 > \alpha_3 w) - P(\alpha_1 X_1 > \alpha_1 u)P(\alpha_2 X_2 > \alpha_2 v)P(\alpha_3 X_3 > \alpha_3 w)$. According to Theorem 1 from Nelsen and Úbeda-Flores [20], $\rho_3^{(\alpha_1,\alpha_2,\alpha_3)}(C)$ is a linear combination of the pairwise measures and the measures $\rho_3^+$ and $\rho_3^-$, given by

$$\rho_3^{(\alpha_1,\alpha_2,\alpha_3)} = \frac{\alpha_1\alpha_2\rho_{12} + \alpha_1\alpha_3\rho_{13} + \alpha_2\alpha_3\rho_{23}}{3} + \alpha_1\alpha_2\alpha_3 \frac{(\rho_3^+ - \rho_3^-)}{2}. \tag{6}$$

Equivalently, $\rho_3^{(\alpha_1,\alpha_2,\alpha_3)}(C)$ is equal to $\rho_3^+(C')$, where $C'$ is the copula associated with the random variables $(\alpha_1 X_1, \alpha_2 X_2, \alpha_3 X_3)$. The purpose of the directional $\rho$-coefficients $\rho_3^{(\alpha_1,\alpha_2,\alpha_3)}$ is to detect positive dependence among the random variables $X_1, X_2, X_3$ undetected by the coefficients $\rho_3^*$, $\rho_3^+$ and $\rho_3^-$. For example, if $(X_1, X_2, X_3)$ are Unif(0, 1) random

**Table 1**
Direction of maximal dependence.

| $\max\{\rho_{12}, \rho_{13}, \rho_{23}, 3\rho_3^*\}$ | $\rho_3^+ - \rho_3^-$ | $(\alpha_1, \alpha_2, \alpha_3)$ |
|---|---|---|
| $3\rho_3^*$ | $\neq 0$ | $\alpha_1 = \alpha_2 = \alpha_3 = sgn(\rho_3^+ - \rho_3^-)$ |
| $3\rho_3^*$ | $= 0$ | $\alpha_1 = \alpha_2 = \alpha_3 = \pm 1$ |
| $\rho_{ij}$ | $\neq 0$ | $-\alpha_i = -\alpha_j = \alpha_k = sgn(\rho_3^+ - \rho_3^-)$ |
| $\rho_{ij}$ | $= 0$ | $-\alpha_i = -\alpha_j = \alpha_k = \pm 1$ |

variables whose joint distribution function is the 3-copula $C(u, v, w) = C_1(\min(u, v), w)$, where $C_1$ is the 2-copula given by $C_1(u, v) = \frac{1}{2}[uv + \max(u + v - 1, 0)]$, then $\rho_3^* = \rho_3^+ = \rho_3^- = 0$. However, there is positive dependence undetected by these coefficients since $P(X_1 = X_2 = 1 - X_3) = \frac{1}{2}$, i.e., half the probability mass is uniformly distributed in the unit cube $[0, 1]^3$ on the line segment joining the points (0, 0, 1) and (1, 1, 0). This positive dependence is detected by the directional $\rho$-coefficients $\rho_3^{(-1,-1,1)} = \rho_3^{(1,1,-1)} = \frac{2}{3}$. The direction $(-1, -1, 1)$ refers to the direction of the inequalities $X_1 \le u, X_2 \le v, X_3 > w$ used in the computation of $\rho_3^{(-1,-1,1)}$. This can be interpreted as "small values of $X_1$ and $X_2$ tend to occur with large values of $X_3$", or roughly that probability is concentrated in the portion of the unit cube $[0, 1]^3$ near the vertex (0, 0, 1). The measures $\rho_3^+$, $\rho_3^-$, and $\rho_3^*$ only measure dependence in the directions (1, 1, 1) and $(-1, -1, -1)$. In the next section we will define an index of positive dependence in trivariate distributions based on the largest of the eight directional $\rho$-coefficients given by Eq. (6).

## 3. New index of positive dependence

**Definition 3.1.** Let $(X_1, X_2, X_3)$ be a random vector with associated 3-copula $C$. Let $\rho_3^{(\alpha_1,\alpha_2,\alpha_3)}(C)$ denote the coefficient of directional dependence given by Eq. (5), with $\alpha_i \in \{-1, 1\}$. Then the index of maximal dependence is given by

$$\rho_3^{\max}(C) = \max_{(\alpha_1,\alpha_2,\alpha_3)} \left\{\rho_3^{(\alpha_1,\alpha_2,\alpha_3)}(C)\right\}.$$

**Theorem 3.1.** *Let $(X_1, X_2, X_3)$ be a random vector with associated 3-copula $C$. Then*

$$\rho_3^{\max} = \frac{2}{3}\max\left\{\rho_{12}, \rho_{13}, \rho_{23}, 3\rho_3^*\right\} - \min\left\{\rho_3^+, \rho_3^-\right\}, \tag{7}$$

*where $\rho_3^*$, $\rho_3^-$ and $\rho_3^+$ are given by Eqs. (2)–(4) respectively.*

**Proof.** According to the relations among $\rho_3^+$, $\rho_3^-$, $\rho_3^*$ and the pairwise measures $\rho_{ij}$, $i \neq j, i, j = 1, 2, 3$, explored in Nelsen and Úbeda-Flores [20], the eight possible cases of Eq. (6) are

$$\rho_3^{(1,1,1)} = 2\rho_3^* - \rho_3^-, \qquad \rho_3^{(-1,-1,-1)} = 2\rho_3^* - \rho_3^+,$$
$$\rho_3^{(-1,-1,1)} = \frac{2}{3}\rho_{12} - \rho_3^-, \qquad \rho_3^{(1,1,-1)} = \frac{2}{3}\rho_{12} - \rho_3^+,$$
$$\rho_3^{(-1,1,-1)} = \frac{2}{3}\rho_{13} - \rho_3^-, \qquad \rho_3^{(1,-1,1)} = \frac{2}{3}\rho_{13} - \rho_3^+,$$
$$\rho_3^{(1,-1,-1)} = \frac{2}{3}\rho_{23} - \rho_3^-, \qquad \rho_3^{(-1,1,1)} = \frac{2}{3}\rho_{23} - \rho_3^+,$$

from which equation (7) follows. □

To determine the direction $(\alpha_1, \alpha_2, \alpha_3)$ which produces the maximal value of $\rho_3^{(\alpha_1,\alpha_2,\alpha_3)}$ we consider conditions about the values of $\max\left\{\rho_{12}, \rho_{13}, \rho_{23}, 3\rho_3^*\right\}$ and $\rho_3^+ - \rho_3^-$, as given in Table 1, where *sgn* denotes the signum function.

Table 1 leads to the following observations.

1. We say that there exists positive dependence undetected by $\rho_3^+$ or $\rho_3^-$ whenever $\rho_3^{\max}$ is not equal to either $\rho_3^+$ or $\rho_3^-$.
2. If $\rho_{12}$, $\rho_{23}$ and $\rho_{13}$ are all positive, then $\rho_3^{\max}$ is equal to either $\rho_3^+$ or $\rho_3^-$, i.e., there is no undetected positive dependence.
3. If at least two of $\rho_{12}$, $\rho_{23}$ and $\rho_{13}$ are negative, then $\rho_3^{\max}$ is not equal to either $\rho_3^+$ or $\rho_3^-$, i.e., there is undetected positive dependence.
4. If exactly one of $\rho_{12}$, $\rho_{23}$ and $\rho_{13}$ is negative, then, there is undetected positive dependence if and only if the sum of the smaller two of $\{\rho_{12}, \rho_{23}, \rho_{13}\}$ is negative.

**Example 3.1.** Let $C$ be the copula that distributes probability mass uniformly on the three line segments in $[0, 1]^3$ joining the point $(\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$ to the vertices $(0, 1, 1)$, $(1, 0, 1)$, and $(1, 1, 0)$. Here $\rho_{12} = \rho_{23} = \rho_{13} = -\frac{1}{3}$ and $\rho_3^* = -\frac{1}{3}$, $\rho_3^- = -\frac{1}{9}$, $\rho_3^+ = -\frac{5}{9}$. As a consequence, $\rho_3^{\max} = \frac{1}{3}$. The directions given by Table 1 are $(1, 1, -1)$ or $(1, -1, 1)$ or $(-1, 1, 1)$ since $\max\left\{\rho_{12}, \rho_{13}, \rho_{23}, 3\rho_3^*\right\} = \rho_{12} = \rho_{13} = \rho_{23}$ and $sgn(\rho_3^+ - \rho_3^-) = -1$. All the traditional measures $\rho_{12}, \rho_{13}, \rho_{23}, \rho_3^*, \rho_3^+, \rho_3^-$ are negative, but $\rho_3^{\max}$ is positive. This means that the index finds positive dependence undetected by $\rho_3^*$, $\rho_3^+$ and $\rho_3^-$. Note that $\rho_3^{(1,1,-1)} = \frac{1}{3}$ indicates that "large" values of $X_1$ and $X_2$ tend to occur with "small" values of $X_3$, but $\rho_3^{(-1,-1,1)} = -\frac{1}{9}$ indicates that it is not the case that the complementary case holds, i.e., that "small" values of $X_1$ and $X_2$ tend to occur with "large" values of $X_3$.

### 3.1. Indexes based on Kendall's tau and Blomqvist's beta

There are also directional coefficients based on the three dimensional versions of the population versions of the measures of association known as Kendall's tau and Blomqvist's beta studied in Nelsen and Úbeda-Flores [20]:

$$\tau_3^{(\alpha_1,\alpha_2,\alpha_3)} = \frac{\alpha_1\alpha_2\tau_{12} + \alpha_1\alpha_3\tau_{13} + \alpha_2\alpha_3\tau_{23}}{3}$$

and

$$\beta_3^{(\alpha_1,\alpha_2,\alpha_3)} = \frac{\alpha_1\alpha_2\beta_{12} + \alpha_1\alpha_3\beta_{13} + \alpha_2\alpha_3\beta_{23}}{3}.$$

These coefficients lead to indexes of maximal dependence similar to $\rho_3^{\max}(C)$:

$$\tau_3^{\max}(C) = \max_{(\alpha_1,\alpha_2,\alpha_3)} \left\{\tau_3^{(\alpha_1,\alpha_2,\alpha_3)}(C)\right\}$$

and

$$\beta_3^{\max}(C) = \max_{(\alpha_1,\alpha_2,\alpha_3)} \left\{\beta_3^{(\alpha_1,\alpha_2,\alpha_3)}(C)\right\}.$$

However, since these indexes do not incorporate a measure of mutual dependence among the three random variables $X_1$, $X_2$ and $X_3$ analogous to $\rho_3^-$ and $\rho_3^+$, they are not as effective in detecting positive dependence. As an example, for the copula in Example 3.1 we have $\tau_3^{\max} = \frac{1}{9}$ occurring in 6 directions (all except $(1, 1, 1)$ and $(-1, -1, -1)$) and $\beta_3^{\max} = 0$ in all 8 directions. Hence in the sequel we will restrict our study to properties of $\rho_3^{\max}$.

### 3.2. Properties of $\rho_3^{\max}$

In this section we present some properties of the index $\rho_3^{\max}$. For a vector $(X_1, X_2, X_3)$ of continuous random variables with copula $C$, we will write both $\rho_3^{\max}(C)$ and $\rho_3^{\max}(X_1, X_2, X_3)$ for the index.

**Theorem 3.2.** *Under the assumptions of Definition 3.1 and the hypotheses of Theorem 3.1, we have the following.*

(i) *The index $\rho_3^{\max}$ is well-defined.*
(ii) $0 \le \rho_3^{\max} \le 1$, *and if $\rho_3^{\max} = 0$, then $\rho_3^{(\alpha_1,\alpha_2,\alpha_3)} = 0$ for every direction $(\alpha_1, \alpha_2, \alpha_3)$ and $\rho_{12} = \rho_{23} = \rho_{13} = \rho_3^* = \rho_3^- = \rho_3^+ = 0$. $\rho_3^{\max}(C_1) = 0$ and $\rho_3^{\max}(C_2) = 1$, where $C_1(u, v, w) = uvw$ and $C_2(u, v, w) = \min\{u, v, w\}$.*
(iii) *$\rho_3^{\max}$ is invariant under permutations, that is, if $\pi$ is a permutation of $\{1, 2, 3\}$, then $\rho_3^{\max}(X_1, X_2, X_3) = \rho_3^{\max}(X_{\pi(1)}, X_{\pi(2)}, X_{\pi(3)})$.*
(iv) *$\rho_3^{\max}$ is invariant under monotone transformations, that is, if $T_1$ is a strictly increasing or strictly decreasing function of $X_1$, then $\rho_3^{\max}(X_1, X_2, X_3) = \rho_3^{\max}(T_1(X_1), X_2, X_3)$ and similarly for $T_2(X_2)$ and $T_3(X_3)$.*
(v) *$\rho_3^{\max}$ is continuous in the following sense: if $\lim_{k\to\infty} C_k = C$ (point wise) for all $u, v, w \in [0, 1]$, then $\lim_{k\to\infty} \rho_3^{\max}(C_k) = \rho_3^{\max}(C)$.*

**Proof.** (i) When the random variables are continuous, the copula of $(X_1, X_2, X_3)$ is unique.

(ii) Since $\sum_{(\alpha_1,\alpha_2,\alpha_3)} \rho_3^{(\alpha_1,\alpha_2,\alpha_3)} = 0$ (see Nelsen and Úbeda-Flores [20] for a proof), the assumption that $\rho_3^{\max} < 0$ leads to a contradiction, hence $\rho_3^{\max} \ge 0$. Since $\rho_3^{(\alpha_1,\alpha_2,\alpha_3)} \le 1$ for every direction $(\alpha_1, \alpha_2, \alpha_3)$, it follows that $\rho_3^{\max} \le 1$. The consequences of $\rho_3^{\max} = 0$ derive from the 8 equations in the proof of Theorem 3.1. But $\rho_3^{\max} = 0$ does not imply that $X_1, X_2, X_3$ are pairwise or mutually independent.

(iii) When $\pi$ is a permutation of $\{1, 2, 3\}$, we have $\rho_3^{(\alpha_1,\alpha_2,\alpha_3)}(X_1, X_2, X_3) = \rho_3^{(\alpha_{\pi(1)},\alpha_{\pi(2)},\alpha_{\pi(3)})}(X_{\pi(1)}, X_{\pi(2)}, X_{\pi(3)})$, from which the result follows.

(iv) If $T_1$ is a strictly increasing transformation, then $\rho_3^{(\alpha_1,\alpha_2,\alpha_3)}(X_1, X_2, X_3) = \rho_3^{(\alpha_1,\alpha_2,\alpha_3)}(T_1(X_1), X_2, X_3)$, and if $T_1$ is a strictly decreasing transformation, then $\rho_3^{(\alpha_1,\alpha_2,\alpha_3)}(X_1, X_2, X_3) = \rho_3^{(-\alpha_1,\alpha_2,\alpha_3)}(T_1(X_1), X_2, X_3)$, from which the result follows.

(v) The integrand $Q_{(\alpha_1,\alpha_2,\alpha_3)}$ in Eq. (5) is a difference of two copulas, and copulas are uniformly continuous on their domain, which is sufficient to establish the result. □

However, the index $\rho_3^{\max}$ is not a measure of multivariate concordance as defined by Taylor [26,27] and Dolati and Úbeda-Flores [5], as it does not satisfy the property of monotonicity. A copula-based measure $\mu$ is *monotone* if $C_1 \prec C_2$ implies $\mu(C_1) \leq \mu(C_2)$, and that is not the case for $\rho_3^{\max}$. For a counterexample, let $C_1(u, v, w) = \max(\min(u, v) + w - 1, 0)$ and $C_2(u, v, w) = w \min(u, v)$. Then $C_1 \prec C_2$, however $\rho_3^{\max}(C_1) = 1 > \frac{1}{3} = \rho_3^{\max}(C_2)$.

### 3.3. Extensions of the index $\rho_3^{\max}$ for dimension $\geq 4$

In theory our work can be extended to $d$-dimensional vectors of continuous random variables. If $C_d$ denotes the $d$-dimensional copula associated with such a vector, then the $d$-dimensional versions of (3) and (4) are given by (Joe [17], Nelsen [19])

$$\rho_d^-(C_d) = \frac{d+1}{2^d - (d+1)}\left(2^d \int_{I^d} C_d(\mathbf{u})d\mathbf{u} - 1\right), \tag{8}$$

$$\rho_d^+(C_d) = \frac{d+1}{2^d - (d+1)}\left(2^d \int_{I^d} \overline{C}_d(\mathbf{u})d\mathbf{u} - 1\right), \tag{9}$$

where $\mathbf{u} = (u_1, \ldots, u_d)$ is a $d$-dimensional vector and $\overline{C}_d$ is the survival function associated with $C_d$. It is natural to extend the definitions of $\rho_3^{\alpha}$ and $\rho_3^{\max}$ as follows:

$$\rho_d^{(\alpha_1,\ldots,\alpha_d)}(C_d) = \frac{d+1}{2^d - (d+1)} \int_{I^d} Q_{\alpha_1,\ldots,\alpha_d}(\mathbf{u})d\mathbf{u}, \tag{10}$$

where $Q_{\alpha_1,\ldots,\alpha_d}(\mathbf{u})$ is $P(\alpha_i X_i > \alpha_i u_i;\, \alpha_i, i = 1, \ldots, d) - \prod_{i=1}^d P(\alpha_i X_i > \alpha_i u_i)$ and

$$\rho_d^{\max}(C_d) = \max_{(\alpha_1,\ldots,\alpha_d)} \left\{\rho_d^{(\alpha_1,\ldots,\alpha_d)}(C_d)\right\}.$$

For $d \geq 4$ the $2^d$ directional coefficients $\rho_d^{(\alpha_1,\ldots,\alpha_d)}(C_d)$ and $\rho_d^{\max}(C_d)$ are then functions of $\binom{d}{2}$ pairwise Spearman's rho coefficients and the $k$-wise coefficients $\rho_k^+(C_k)$ and $\rho_k^-(C_k)$ for $3 \leq k \leq d$, where $C_k$ (for $3 \leq k \leq d$) denotes a $k$-dimensional margin of $C_d$.

The complexity in evaluating $\rho_d^{\max}(C_d)$ from $d$-dimensional versions of Theorem 3.1 grows exponentially in $d$. For example, when $d = 4$ the 16 directional coefficients are functions of 16 pairwise and $k$-wise versions of Spearman's rho; when $d = 5$ the 32 directional coefficients are functions of 42 pairwise and $k$-wise versions of Spearman's rho; and when $d = 6$ the 64 directional coefficients are functions of 99 pairwise and $k$-wise versions of Spearman's rho. The index $\rho_d^{\max}$ for $d \geq 4$ awaits further study.

## 4. Estimators

Consider a trivariate random sample $\left\{(X_{1j}, X_{2j}, X_{3j})\right\}_{j=1}^n$ of the vector $(X_1, X_2, X_3)$ with associated unknown copula $C$. Let be $R_{ij}$ = rank of $X_{ij}$ in $\{X_{i1}, \ldots, X_{in}\}$ and define $\overline{R}_{ij} = n + 1 - R_{ij}$, for $i = 1, 2, 3$. The nonparametric estimators of each coefficient (given by equations (1) and (3)) are well-known (see Joe [17]) and they are respectively given by

$$\hat{\rho}_{ik} = \frac{12}{n(n^2-1)} \sum_{j=1}^n R_{ij}R_{kj} - 3\frac{(n+1)}{(n-1)}, \quad ik \in \{12, 23, 13\} \tag{11}$$

$$\hat{\rho}_3^- = \frac{8}{n(n-1)(n+1)^2} \sum_{j=1}^n R_{1j}R_{2j}R_{3j} - \frac{(n+1)}{(n-1)}. \tag{12}$$

It is easy to derive the estimator of $\rho_3^+$ from Eq. (12)

$$\hat{\rho}_3^+ = \frac{8}{n(n-1)(n+1)^2} \sum_{j=1}^n \overline{R}_{1j}\overline{R}_{2j}\overline{R}_{3j} - \frac{(n+1)}{(n-1)}. \tag{13}$$

In the next definition we introduce estimators of each $\rho_3^{(\alpha_1,\alpha_2,\alpha_3)}$.

**Definition 4.1.** Define $R_{ij}^{\alpha_i}$ to be $R_{ij}$ if $\alpha_i = -1$ and $\overline{R}_{ij}$ if $\alpha_i = 1$, and set

$$\hat{\rho}_3^{(\alpha_1,\alpha_2,\alpha_3)} = \frac{8}{n(n-1)(n+1)^2} \sum_{j=1}^n R_{1j}^{\alpha_1}R_{2j}^{\alpha_2}R_{3j}^{\alpha_3} - \frac{(n+1)}{(n-1)}.$$

**Remark 4.1.** $\hat{\rho}_3^{(\alpha_1,\alpha_2,\alpha_3)}$ estimates $\rho_3^{(\alpha_1,\alpha_2,\alpha_3)}$. For example, when $(\alpha_1,\alpha_2,\alpha_3)=(-1,-1,1)$ we have

$$\begin{aligned}\hat{\rho}_3^{(-1,-1,1)} &= \frac{8}{n(n-1)(n+1)^2}\sum_{j=1}^{n} R_{1j}R_{2j}(n+1-R_{3j}) - \frac{(n+1)}{(n-1)}\\ &= \frac{8}{n(n^2-1)}\sum_{j=1}^{n} R_{1j}R_{2j} - 2\frac{(n+1)}{(n-1)} - \frac{8}{n(n-1)(n+1)^2}\sum_{j=1}^{n} R_{1j}R_{2j}R_{3j} + \frac{(n+1)}{(n-1)}\\ &= \frac{2}{3}\hat{\rho}_{12} - \hat{\rho}_3^-\end{aligned}$$

which estimates $\rho_3^{(-1,-1,1)}$ since $\rho_3^{(-1,-1,1)} = \frac{2}{3}\rho_{12} - \rho_3^-$ (see the relations used in the proof of Theorem 3.1), and the other seven cases are similar.

As a consequence of Theorem 3.1, Definition 4.1 and Remark 4.1 we introduce the next estimator.

**Definition 4.2.** The plug-in estimator of $\rho_3^{\max}$ is

$$\hat{\rho}_3^{\max} = \frac{2}{3}\max\left\{\hat{\rho}_{12}, \hat{\rho}_{13}, \hat{\rho}_{23}, 3\hat{\rho}_3^*\right\} - \min\left\{\hat{\rho}_3^+, \hat{\rho}_3^-\right\},$$

where $3\hat{\rho}_3^* = \hat{\rho}_{12} + \hat{\rho}_{13} + \hat{\rho}_{23}$.

**Remark 4.2.** $\hat{\rho}_3^{\max} = \max_\alpha\left\{\hat{\rho}_3^\alpha\right\}$. Given each direction $\alpha$, we can show using Remark 4.1 that the estimator $\hat{\rho}_3^\alpha$ of $\rho_3^\alpha$ follows one of the 8 equations exhibited in the proof of Theorem 3.1, replacing $\rho_{ik}, ik \in \{12, 13, 23\}\, \rho_3^+, \rho_3^-$ and $\rho_3^*$ by $\hat{\rho}_{ik}, ik \in \{12, 13, 23\}, \hat{\rho}_3^+, \hat{\rho}_3^-$ and $\hat{\rho}_3^*$ respectively, then by the same arguments used to prove Theorem 3.1 $\max_\alpha\left\{\hat{\rho}_3^\alpha\right\}$ is given by Definition 4.2.

## 5. Empirical processes related to $\rho_{ij}$, $\rho_3^\alpha$ and $\rho_3^{\max}$

Let ß be an index set, such that ß $\subseteq \{1, 2, 3\}$. We define, for ß $= \{1, 2, 3\}$, $x_ß = (x_1, x_2, x_3)$ an arbitrary value of $(X_1, X_2, X_3)$; for ß $= \{i, k\}$, $x_ß = (x_i, x_k)$ an arbitrary value of $(X_i, X_k)$. Let $|ß|$ denote the cardinal of ß and $\alpha = (\alpha_1, \alpha_2, \alpha_3)$. Consider the function,

$$H_{ß,\alpha}(x_ß) = P(\alpha_i X_i \le \alpha_i x_i, i \in ß) \tag{14}$$

called here simply the $|ß|$-dimensional distribution function. Let $u_ß = (u_1, u_2, u_3)$ for ß $= \{1, 2, 3\}$, $u_ß = (u_i, u_k)$ for ß $= \{i, k\}$ and $F_i$ the marginal cumulative distribution function of $X_i$. Let $F_i^{-1}$ denote the inverse of $F_i$, $i = 1, 2, 3$

$$C_{ß,\alpha}(u_ß) = H_{ß,\alpha}(F_i^{-1}(u_i), \quad i \in ß) \tag{15}$$

which is a generalization of a 3-copula when $\alpha_i = 1, i \in ß = \{1, 2, 3\}$.

We introduce the empirical process to estimate the previous function

$$C_{ß,\alpha,n}(u_ß) = \frac{1}{n+1}\sum_{j=1}^{n}\prod_{i\in ß} 1_{\left\{\alpha_i \frac{R_{ij}}{n+1} \le \alpha_i u_i\right\}}, \tag{16}$$

where (16) gives the empirical estimator of the copula, when ß $= \{1, 2, 3\}$ and $\alpha = (1, 1, 1)$.

**Remark 5.1.** If we define the estimators

$$H_{ß,\alpha,n}(x_ß) = \frac{1}{n+1}\sum_{j=1}^{n}\prod_{i\in ß} 1_{\{\alpha_i X_{ij} \le \alpha_i x_i\}} \tag{17}$$

$$F_{i,n}(x) = \frac{1}{n+1}\sum_{j=1}^{n} 1_{\{X_{ij} \le x\}} \tag{18}$$

where $F_{i,n}(x_{ij}) = \frac{R_{ij}}{n+1}$ and $x_{ij}$ is the observed value of $X_{ij}, j = 1, \ldots, n, i \in ß$, and let

$$F_{i,n}^{-1}(u) = \inf\left\{x \in \mathbb{R} : F_{i,n}(x) \ge u\right\}, \quad u \in [0, 1], \tag{19}$$

then, we obtain $C_{ß,\alpha,n}(u_ß) = H_{ß,\alpha,n}(F_{i,n}^{-1}(u_i), \ i \in ß)$.

In order to derive the weak convergence of the empirical processes

$$\sqrt{n}\left\{C_{ß,\alpha,n}(u_ß) - C_{ß,\alpha}(u_ß)\right\}, \quad u_ß \in [0,1]^{|ß|} \tag{20}$$

we introduce a condition on $C_{ß,\alpha}$ inspired by Segers [25].

For $i \in ß$ if $e_i$ is a vector such that $(e_i)_j = 0, j \neq i$, $(e_i)_j = 1$, $j = i, j \in ß$, define the $i$-th first-order partial derivative of $C_{ß,\alpha}$, as

$$\dot{C}_{i,ß,\alpha}(u_ß) = \lim_{h\to 0} \frac{C_{ß,\alpha}(u_ß + he_i) - C_{ß,\alpha}(u_ß)}{h} \quad \text{for } u_ß \in [0,1]^{|ß|}.$$

**Condition 5.1.** For each $i \in ß$, the $i$-th first-order partial derivative $\dot{C}_{i,ß,\alpha}$ exists and is continuous on the set $\{u_ß \in [0,1]^{|ß|} : 0 < u_i < 1\}$.

We also extend the function $\dot{C}_{i,ß,\alpha}$ to the boundary as follows. If $u_ß \in [0,1]^{|ß|}$ and $u_i = 0$, $\dot{C}_{i,ß,\alpha}(u_ß) = \limsup_{h\downarrow 0} \frac{C_{ß,\alpha}(u_ß+he_i)-C_{ß,\alpha}(u_ß)}{h}$. If $u_ß \in [0,1]^{|ß|}$ and $u_i = 1$, $\dot{C}_{i,ß,\alpha}(u_ß) = \limsup_{h\downarrow 0} \frac{C_{ß,\alpha}(u_ß)-C_{ß,\alpha}(u_ß-he_i)}{h}$.

The next theorem is valid for dimensions $d > 3$. Nevertheless for our purpose $d \leq 3$ suffices. In this theorem we will show that the empirical process, given by (20) goes weakly to

$$\mathbb{G}_{C_{ß,\alpha}}(u_ß) = \mathbb{B}_{C_{ß,\alpha}}(u_ß) - \sum_{i\in ß} \dot{C}_{i,ß,\alpha}(u_ß)\mathbb{B}_{C_{ß,\alpha}}(u_ß^{(i)}), \tag{21}$$

where $\mathbb{G}_{C_{ß,\alpha}}(u_ß)$ follows the next condition.

**Condition 5.2.** $\mathbb{B}_{C_{ß,\alpha}}(u_ß)$ is a $C_{ß,\alpha}$-tight centered Gaussian process on $[0,1]^{|ß|}$, $u_ß^{(i)}$ is a vector such that the j-th component, $j \in ß$, $(u_ß^{(i)})_j = 1$ if $j \neq i$ when $\alpha_j = 1$, $(u_ß^{(i)})_j = 0$ if $j \neq i$ when $\alpha_j = -1$, and for $j = i$, $(u_ß^{(i)})_i = u_i$, $i \in ß$. The covariance function is $\mathbb{E}(\mathbb{B}_{C_{ß,\alpha}}(u_ß)\mathbb{B}_{C_{ß,\alpha}}(v_ß)) = C_{ß,\alpha}(w_ß) - C_{ß,\alpha}(u_ß)C_{ß,\alpha}(v_ß)$, where the j-th component $(w_ß)_j = u_j \wedge v_j$ if $\alpha_j = 1$ and $(w_ß)_j = u_j \vee v_j$ if $\alpha_j = -1$, $j \in ß$.

**Theorem 5.1.** *Let $H_{ß,\alpha}$ be a $|ß|$-dimensional distribution function, given by Eq. (14) with continuous marginal distributions $F_i$, $i \in ß$ and with $C_{ß,\alpha}$ given by Eq. (15), where $ß \subseteq \{1,2,3\}$ and $\alpha = (\alpha_1, \alpha_2, \alpha_3)$, $\alpha_i \in \{-1, 1\}$. Under the additional Condition 5.1 on the function $C_{ß,\alpha}$, when $n \to \infty$*

$$\sqrt{n}\left\{C_{ß,\alpha,n}(u_ß) - C_{ß,\alpha}(u_ß)\right\} \to^{w} \mathbb{G}_{C_{ß,\alpha}}(u_ß). \tag{22}$$

*Weak convergence takes place in $\ell^\infty([0,1]^{|ß|})$ and $\mathbb{G}_{C_{ß,\alpha}}(u_ß) = \mathbb{B}_{C_{ß,\alpha}}(u_ß) - \sum_{i\in ß} \dot{C}_{i,ß,\alpha}(u_ß)\mathbb{B}_{C_{ß,\alpha}}(u_ß^{(i)})$, where $\mathbb{B}_{C_{ß,\alpha}}(u_ß)$ is a $C_{ß,\alpha}$-tight centered Gaussian process on $[0,1]^{|ß|}$ and $\mathbb{G}_{C_{ß,\alpha}}(u_ß)$ follows Condition 5.2.*

**Proof.** Consider the empirical process $\mathbb{B}_{n,C_{ß,\alpha}}(u_ß) = \sqrt{n}(G_{ß,\alpha,n}(u_ß) - C_{ß,\alpha}(u_ß))$ where for $U_{ij} = F_i(X_{ij})$, $i \in ß, j = 1, \ldots, n$ and $u_ß \in [0,1]^{|ß|}$,

$$G_{ß,\alpha,n}(u_ß) = \frac{1}{n+1} \sum_{j=1}^{n} \prod_{i\in ß} 1_{\{\alpha_i U_{ij} \leq \alpha_i u_i\}}. \tag{23}$$

Note that if $ß = \{1,2,3\}$, $\alpha = (1,-1,1)$, $i = 3$ by hypothesis $u_ß^{(i)} = (1, 0, u_3)$ $\mathbb{B}_{n,C_{ß,\alpha}}(u_ß^{(i)}) = \sqrt{n}\Big(\frac{1}{n+1}\sum_{j=1}^{n} 1_{\{U_{3j}\leq u_3\}} - P(U_3 \leq u_3)\Big) \to_{n\to\infty} 0$, for $u_3 = 0$ and $u_3 = 1$. Using the same arguments, for arbitrary $i$, $\mathbb{B}_{n,C_{ß,\alpha}}(u_ß^{(i)}) \to_{n\to\infty} 0$ on the boundary $u_i = 0$ and $u_i = 1$.

Define the process

$$\tilde{\mathbb{G}}_{C_{ß,\alpha}}(u_ß) = \mathbb{B}_{n,C_{ß,\alpha}}(u_ß) - \sum_{i\in ß} \dot{C}_{i,ß,\alpha}(u_ß)\mathbb{B}_{n,C_{ß,\alpha}}(u_ß^{(i)}). \tag{24}$$

The process $\mathbb{B}_{C_{ß,\alpha}}$ is the weak limit in $\ell^\infty([0,1]^{|ß|})$ of the sequence $\left\{\mathbb{B}_{n,C_{ß,\alpha}}\right\}_{n\geq 1}$, where $\mathbb{B}_{C_{ß,\alpha}}$ is a $C_{ß,\alpha}$-Brownian bridge and it can be assumed to have continuous trajectories (by the Empirical Central Limit Theorem, see Van der Vaart and Wellner [28]).

From Condition 5.1 and assuming the extension of the partial derivatives to the whole of $[0,1]^{|ß|}$, and that the trajectories of $\mathbb{B}_{C_{ß,\alpha}}$ are continuous, the trajectories of $\mathbb{G}_{C_{ß,\alpha}}$ are also continuous. In fact, when $\dot{C}_{i,ß,\alpha}$ fail to be continuous for $u_ß \in [0,1]^{|ß|}$ such that $u_i = 0$ or $u_i = 1$ we have $\mathbb{B}_{C_{ß,\alpha}}(u_ß^{(i)}) = 0$ also. The process $\mathbb{G}_{C_{ß,\alpha}}$ is the weak limit in $\ell^\infty([0,1]^{|ß|})$ of the sequence $\left\{\tilde{\mathbb{G}}_{n,C_{ß,\alpha}}\right\}_{n\geq 1}$.

If Condition 5.1 holds, following the same arguments in the proof of Proposition 3.1 in Segers [25], where Condition 5.1 is used to apply the mean value theorem over $C_{ß,\alpha}$, convergence (in probability) follows

$$\sup_{u_ß\in[0,1]^{|ß|}} |\sqrt{n}\left\{C_{ß,\alpha,n} - C_{ß,\alpha}\right\} - \tilde{\mathbb{G}}_{n,C_{ß,\alpha}}| \to^{P} 0, \text{ when } n\to\infty.$$

Then, the weak convergence stated by Eq. (22) also follows.

The covariance function is derived applying the multidimensional Central Limit Theorem; see for example Gänßler [9]. □

A special case of the previous theorem is proved in Schmid and Schmidt [24] (Theorem 2, page 411), assuming an arbitrary dimension and some additional conditions for the joint cumulative distribution.

We observe, for each $i$ and $j$,

$$\int_I 1_{\left\{\alpha_i \frac{R_{ij}}{n+1}\le \alpha_i u_i\right\}} du_i = \frac{R_{ij}^{\alpha_i}}{n+1}, \tag{25}$$

and as a consequence, according to Eq. (16),

$$\int_{I^{|ß|}} C_{ß,\alpha,n}(u_ß)du_ß = \frac{1}{(n+1)^{|ß|+1}}\sum_{j=1}^{n}\prod_{i\in ß} R_{ij}^{\alpha_i}. \tag{26}$$

### 5.1. Properties of estimators

This subsection explores the relationships between empirical processes and the pairwise Spearman's rho coefficients and coefficients of directional dependence.

**Remark 5.2.** Using Eqs. (11)–(13) and Definition 4.1 we obtain,

(i) $\hat{\rho}_{ik} = 12\frac{(n+1)^2}{n(n-1)}\int_{I^2} C_{ß,\alpha,n}(u_ß)du_ß - 3\frac{(n+1)}{(n-1)}$, with $ß=\{i,k\}$, $\alpha_i=\alpha_k=1$, and $ik\in\{12,23,13\}$;

(ii) $\hat{\rho}_3^{\alpha} = 8\frac{(n+1)^2}{n(n-1)}\int_{I^3} C_{ß,\alpha,n}(u_ß)du_ß - \frac{(n+1)}{(n-1)}$, with $ß=\{1,2,3\}$, and an arbitrary vector $\alpha$,
- (ii1) If $\alpha_i = 1\,\forall i\in ß$, $\hat{\rho}_3^{\alpha} = \hat{\rho}_3^{+}$;
- (ii2) If $\alpha_i = -1\,\forall i\in ß$, $\hat{\rho}_3^{\alpha} = \hat{\rho}_3^{-}$.

In (16) and (17) we constructed empirical processes rescaled by $(n+1)$, by convenience in order to express the estimators in terms of the empirical processes (see Remark 5.2), since we define $\bar{R}_{ij} = n+1-R_{ij}$.

The proof of the next result is an adaptation of Fermanian et al. [7] (Theorem 6, page 854) and Gänßler and Stute [10], page 55.

**Theorem 5.2.** *Under the assumptions of Theorem 5.1, suppose that the real number sequences $\{a_n\}_{n\ge1}$ and $\{b_n\}_{n\ge1}$ satisfy $\sqrt{n}(a_n-a_0) = O(n^{-1/2})$ and $\sqrt{n}(b_n-b_0) = O(n^{-1/2})$, respectively, where $a_0$ and $b_0$ are constant values. Let $T_n(f) = a_n\int_{I^{|ß|}} f(u_ß)du_ß + b_n$, for $n\ge 0$, where $f$ is a $|ß|$-integrable function. Then, when $n\to\infty$,*

$$\sqrt{n}\left\{T_n(C_{ß,\alpha,n}) - T_0(C_{ß,\alpha})\right\} \to^{w} Z_{C_{ß,\alpha}} \sim N(0,\sigma^2_{C_{ß,\alpha}})$$

*with $\sigma^2_{C_{ß,\alpha}} = a_0^2\int_{I^{|ß|}}\int_{I^{|ß|}} \mathbb{E}[\mathbb{G}_{C_{ß,\alpha}}(u_ß)\mathbb{G}_{C_{ß,\alpha}}(v_ß)]du_ß dv_ß$.*

**Proof.**

$$\begin{aligned}
\sqrt{n}\left\{T_n(C_{ß,\alpha,n}) - T_0(C_{ß,\alpha})\right\} &= \sqrt{n}\left\{T_n(C_{ß,\alpha,n}) - T_0(C_{ß,\alpha,n})\right\} + \sqrt{n}\left\{T_0(C_{ß,\alpha,n}) - T_0(C_{ß,\alpha})\right\} \\
&= \sqrt{n}(a_n-a_0)\int_{I^{|ß|}} C_{ß,\alpha,n}(u_ß)du_ß + \sqrt{n}(b_n-b_0) \\
&\quad + a_0\int_{I^{|ß|}} \sqrt{n}\left\{C_{ß,\alpha,n}(u_ß) - C_{ß,\alpha}(u_ß)\right\}du_ß \\
&= a_0\int_{I^{|ß|}} \sqrt{n}\left\{C_{ß,\alpha,n}(u_ß) - C_{ß,\alpha}(u_ß)\right\}du_ß + O(n^{-1/2}) \\
&\to^{w} a_0\int_{I^{|ß|}} \mathbb{G}_{C_{ß,\alpha}}(u_ß)du_ß
\end{aligned} \tag{27}$$

the last equality coming from the assumptions for the sequences $\{a_n\}_{n\geq 1}$ and $\{b_n\}_{n\geq 1}$. The weak convergence follows from Theorem 5.1, using the weak convergence established in Eq. (22), and from Van der Vaart and Wellner [28] (Theorem 1.3.6, in page 20) applied to the continuous integral operator.

A continuous and linear transformation of a tight Gaussian process is normally distributed, so that we can define $Z_{C_{ß,\alpha}} := a_0 \int_{I^{|ß|}} \mathbb{G}_{C_{ß,\alpha}}(u_ß)du_ß$ with distribution $N(0, \sigma^2_{C_{ß,\alpha}})$.

The expression for $\sigma^2_{C_{ß,\alpha}}$ can be obtained by an application of Fubini's theorem. □

**Corollary 5.1.** *Under the assumptions of Theorem 5.2, we have the following.*

1. *When $n \to \infty$, $\sqrt{n}\left\{\hat{\rho}_{ik} - \rho_{ik}\right\} \to^{w} Z_{C_{ß,\alpha}} \sim N(0, \sigma^2_{C_{ß,\alpha}})$, where $ß = \{i, k\}$, the vector $\alpha = (\alpha_1, \alpha_2, \alpha_3)$ is such that $\alpha_i = \alpha_k = 1$, and $ik \in \{12, 23, 13\}$.*
2. *When $n \to \infty$, $\sqrt{n}\left\{\hat{\rho}^{\alpha}_3 - \rho^{\alpha}_3\right\} \to^{w} Z_{C_{ß,\alpha}} \sim N(0, \sigma^2_{C_{ß,\alpha}})$, where $ß = \{1, 2, 3\}$ and the vector $\alpha = (\alpha_1, \alpha_2, \alpha_3)$ is such that $\alpha_i \in \{-1, 1\}$.*

**Proof.** Let $k_0$ and $k_1$ be constants. Given $a_n = k_0 \frac{(n+1)^2}{n(n-1)}$, $a_0 = k_0$, $b_n = -k_1 \frac{(n+1)}{(n-1)}$, $b_0 = -k_1$, the conditions $\sqrt{n}(a_n - a_0) = O(n^{-1/2})$, $\sqrt{n}(b_n - b_0) = O(n^{-1/2})$ are true.

If $k_0 = 12$ and $k_1 = 3$, then Conclusion 1 follows from Remark 5.2(i). Also if $k_0 = 8$ and $k_1 = 1$, then Conclusion 2 follows from Remark 5.2(ii). □

**Theorem 5.3.** *Under the assumptions of Theorem 5.1,*

$$\sqrt{n}\left\{\hat{\rho}^{\max}_3 - \rho^{\max}_3\right\} \to^{w} Z_{C_{ß,\alpha^*}} \sim N(0, \sigma^2_{C_{ß,\alpha^*}}),$$

*where $ß = \{1, 2, 3\}$ and $\alpha^*$ is such that $\rho^{\alpha^*}_3 = \rho^{\max}_3$.*

**Proof.** Define $A^{\alpha}_n = \left\{w : \hat{\rho}^{\alpha^*}_3(w) < \hat{\rho}^{\alpha}_3(w)\right\}$, $\alpha \in \mathcal{A}$ and $A_n = \left\{w : \hat{\rho}^{\alpha^*}_3(w) < \hat{\rho}^{\alpha}_3(w), \forall \alpha \in \mathcal{A}\right\}$, where $\mathcal{A} = \{(\alpha_1, \alpha_2, \alpha_3) : \alpha_i \in \{-1, 1\}, i = 1, 2, 3\}$.

Because $\hat{\rho}^{\alpha}_3 - \hat{\rho}^{\alpha^*}_3 \to \rho^{\alpha}_3 - \rho^{\alpha^*}_3 < 0$ almost surely when $\rho^{\alpha}_3 \neq \rho^{\alpha^*}_3$, then $P(A^{\alpha}_n) \to 0$ when $n \to \infty$ also, because $\hat{\rho}^{\alpha}_3 - \hat{\rho}^{\alpha^*}_3 \to \rho^{\alpha}_3 - \rho^{\alpha^*}_3 = 0$ almost surely when $\rho^{\alpha}_3 = \rho^{\alpha^*}_3$, then $P(A^{\alpha}_n) \to 0$. Hence $P(A_n) \to 0$ when $n \to \infty$.

To show the convergence $\hat{\rho}^{\alpha}_3 - \hat{\rho}^{\alpha^*}_3 \to \rho^{\alpha}_3 - \rho^{\alpha^*}_3$ consider the processes

$$\xi^{\alpha}_n = \sqrt{n}\left\{\hat{\rho}^{\alpha}_3 - \rho^{\alpha}_3\right\}, \quad \alpha \in \mathcal{A}.$$

For each direction $\alpha \in \mathcal{A}$ we can write $\hat{\rho}^{\alpha}_3 - \hat{\rho}^{\alpha^*}_3 = \rho^{\alpha}_3 - \rho^{\alpha^*}_3 + \frac{(\xi^{\alpha}_n - \xi^{\alpha^*}_n)}{\sqrt{n}}$. The difference $\frac{\xi^{\alpha}_n - \xi^{\alpha^*}_n}{\sqrt{n}} \to 0$ almost surely, because the limit variance of the numerator is finite by item (2) of Corollary 5.1.

Consider now

$$\xi^{\max}_n = \sqrt{n}\left\{\max_{\alpha}\left\{\hat{\rho}^{\alpha}_3\right\} - \rho^{\alpha^*}_3\right\}$$

we can establish inferior and superior bounds for the cumulative distribution function of $\xi^{\max}_n$, as follows

$$P(\xi^{\alpha^*}_n \leq x) \leq P(\xi^{\max}_n \leq x) = P(\xi^{\max}_n \leq x, A_n) + P(\xi^{\max}_n \leq x, A^c_n)$$

where the inequality is a consequence of $\hat{\rho}^{\alpha^*}_3 \leq \max_{\alpha}\left\{\hat{\rho}^{\alpha}_3\right\}$.

By the definition of $A_n$, we have $\forall w \in A^c_n$, $\xi^{\max}_n(w) = \xi^{\alpha^*}_n(w)$ almost surely, then

$$P(\xi^{\max}_n \leq x) \leq P(A_n) + P(\xi^{\alpha^*}_n \leq x, A^c_n).$$

As a consequence $P(\xi^{\max}_n \leq x) = P(\xi^{\alpha^*}_n \leq x)$ when $n \to \infty$. By Remark 4.2 and by item (2) of Corollary 5.1 applied over $\hat{\rho}^{\alpha^*}_3$ the result follows. □

**Remark 5.3.** By Theorem 5.3, $\hat{\rho}^{\max}_3$ is an asymptotically unbiased estimator of $\rho^{\alpha^*}_3$ and $\mathrm{Var}(\hat{\rho}^{\max}_3) \to 0$ when $n \to \infty$. As a consequence, by Chebyshev's inequality, we guarantee the convergence in probability, $\hat{\rho}^{\max}_3 \to^{P} \rho^{\alpha^*}_3$ when $n \to \infty$, i.e. $\hat{\rho}^{\max}_3$ is asymptotically consistent.

For an arbitrary dimension $d$ with each component of the vector $\alpha$, $\alpha_i = 1$, $i = 1, \ldots, d$, Deheuvels [4] obtains the decomposition of the process given by Eq. (22) into $2^d - d - 1$ asymptotically independent sub-processes (see Dugué [6]), in order to test for multivariate independence. As summarized in Quessy [21], the same idea holds for an arbitrary dimension $d$, $\alpha_i = -1$, $i = 1 \ldots, d$. The large sample representation of those processes, through the Möbius decomposition of the empirical copula process and of the survival copula process allows us to characterize the asymptotic behavior of five new test statistics, to test independence; see Quessy [21]. It would be natural to investigate, under the conditions of Theorem 5.1 and for an arbitrary value $\alpha_i \in \{-1, 1\}$ how to define a family of statistics to test independence, and obtain its asymptotic distributions and its asymptotic relative efficiency (see Genest et al. [15] and Quessy [21]), those topics await further study.

**Table 2**
Cases simulated. $\mathrm{B}(a, b, c, d)$ and $\mathrm{G}(a, b, c, d)$ denote the trivariate Beta distribution and the trivariate Gamma distribution with parameters $a, b, c, d$ respectively. $\mathrm{B2}(a, b, c, d)$ denotes the distribution of $(Y_1, Y_2, Y_3)$ where $(Y_1, 1 - Y_2, Y_3)$ has distribution $\mathrm{B}(a, b, c, d)$. $D_1D_2D_3(a, b, c)$ denotes the d-vine copula model where $c_{12}$ is the density of a copula $D_1$ with parameter $a$, $c_{23}$ is the density of a copula $D_2$ with parameter $b$ and $c_{13|2}$ is the density of a copula $D_3$ with parameter $c$. $D_i = F$ denotes the Frank copula and $D_i = G$ denotes the Gumbel copula.

| Case | Distribution | Parameters | Observation illustrated |
|---|---|---|---|
| 1 | $\mathrm{G}(a, b, c, d)$ | (1, 0.25, 0.25, 4) | 2 |
| 2 | $\mathrm{GFG}(a, b, c)$ | (5, −7, 2) | |
| 3 | $\mathrm{GGF}(a, b, c)$ | (3, 10, −0.5) | |
| 4 | $\mathrm{FFF}(a, b, c)$ | (−5, −10, −2) | 3 |
| 5 | $\mathrm{B}(a, b, c, d)$ | (1, 2, 2, 4) | |
| 6 | $\mathrm{B}(a, b, c, d)$ | (1, 0.25, 6, 4) | |
| 7 | $\mathrm{B}(a, b, c, d)$ | (1, 4, 0.25, 2) | |
| 8 | $\mathrm{GGF}(a, b, c)$ | (10, 10, −10) | |
| 9 | $\mathrm{FFF}(a, b, c)$ | (−7, −7, −10) | |
| 10 | $\mathrm{B2}(a, b, c, d)$ | (1, 2, 2, 4) | 4 |
| 11 | $\mathrm{B2}(a, b, c, d)$ | (1, 0.25, 4, 4) | |
| 12 | $\mathrm{B2}(a, b, c, d)$ | (1, 4, 0.25, 2) | |

## 6. Simulations and applications

### 6.1. Simulations

We simulated from trivariate Beta and Gamma distributions with diverse parameters. The exact definitions for the models and the simulation methods can be found in Johnson and Kotz [18] (page 231 for the Beta distribution and page 216 for the Gamma distribution). We simulated trivariate d-vine copulas, constructed through combinations of Frank and Gumbel copulas. $\mathrm{B}(a, b, c, d)$ denotes the trivariate Beta distribution with parameters $a, b, c, d$ and $\mathrm{G}(a, b, c, d)$ denotes the trivariate Gamma distribution with parameters $a, b, c, d$. We also simulated random vectors $(X_1, X_2, X_3)$ from the trivariate Beta, $\mathrm{B}(a, b, c, d)$ and we define, $(Y_1, Y_2, Y_3) = (X_1, (1 - X_2), X_3)$. Let $\mathrm{B2}(a, b, c, d)$ denote the distribution of $(Y_1, Y_2, Y_3)$. In the cases of the d-vine copulas, we used the R package "vines" *(Multivariate Dependence Modeling with Vines)* to simulate three different trivariate models with diverse parameters. Following the notation in Aas et al. [1], $D_1D_2D_3(a, b, c)$ denotes the d-vine copula model where $c_{12}$ is the density of a copula $D_1$ with parameter $a$, $c_{23}$ is the density of a copula $D_2$ with parameter $b$ and $c_{13|2}$ is the density of a copula $D_3$ with parameter $c$. For $i = 1, 2, 3$, $D_i = F$ denotes the Frank copula and $D_i = G$ denotes the Gumbel copula.

The accuracy of the estimator $\rho_3^{\max}$ can be estimated using a bootstrap approach (see Schmid and Schmidt [23]). In our simulation study, to evaluate the variance of the $\rho_3^{\max}$ estimator we simulated 1000 samples for each sample size 500, 1000 and 5000.

The choice of Beta and Gamma distributions and the particular structure of the d-vine copulas was made to cover a broad spectrum of the values of the pairwise Spearman's rho and to cover several relationships among the 3-dimensional versions of Spearman's rho. With these we obtain, a variety of directional dependences that show several aspects of the new index. We emphasize that while the copula is used to derive the index, in practice (simulation and data sets), the underlying copula is not needed to estimate the index, we only use the ranks of the observations.

#### 6.1.1. Results

We implemented twelve different cases, given by Table 2, that illustrate the observations following Table 1. For each case and each sample size $n = 500, 1000, 5000$, we show in Table 3, mean values for 1000 simulated samples of $\hat{\rho}_3^+, \hat{\rho}_3^-, \hat{\rho}_3^*, \hat{\rho}_3^{\max}, \hat{\sigma}_{\rho_3^{\max}}$ (standard deviation of $\hat{\rho}_3^{\max}$), mode of the estimated maximal direction $(\hat{\alpha}_1, \hat{\alpha}_2, \hat{\alpha}_3)$ and proportion of times in which the estimated direction was the mode $(\hat{p}_\alpha)$.

Cases 1, 2 and 3 illustrate observation 2. For cases 1 and 2, $\bar{\hat{\rho}}_3^{\max} = \bar{\hat{\rho}}_3^-$ while for case 3 $\bar{\hat{\rho}}_3^{\max} = \bar{\hat{\rho}}_3^+$. Cases 4, 5, 6 and 7 illustrate observation 3, in each case the pairwise correlations and the 3-dimensional versions of Spearman's rho, $\bar{\hat{\rho}}_3^*, \bar{\hat{\rho}}_3^+, \bar{\hat{\rho}}_3^-$ are all negative. Those four cases show the maximal (and positive) $\bar{\hat{\rho}}_3^{\max}$ can be detected in different directions $(\alpha_1, \alpha_2, \alpha_3)$. If we focus on case 7 we note that the scatterplot of the simulated observations (Fig. 1 (left)) shows that the Spearman correlation $\rho_{13}$ is negative but the maximal dependence is not evident. The direction $(-1, 1, 1)$ of maximal dependence is clear from the scatterplot of margins transformed to $[0, 1]$ by scaling ranks; see Fig. 1 (right). We emphasize case 4, in which we illustrate a situation with $\bar{\hat{\rho}}_3^* = \bar{\hat{\rho}}_3^+ = \bar{\hat{\rho}}_3^-$.

Cases 8 and 9 show situations with exactly two negative pairwise correlations (observation 3). In addition, the 3-dimensional versions of Spearman's rho, $\bar{\hat{\rho}}_3^*, \bar{\hat{\rho}}_3^+, \bar{\hat{\rho}}_3^-$ are all negative and take the same value.

**Table 3**
For each case, the mean values of Spearman correlations, $\hat{\rho}_3^+$, $\hat{\rho}_3^-$, $\hat{\rho}_3^*$, $\hat{\rho}_3^{\max}$, $\hat{\sigma}_{\rho_3^{\max}}$ (standard deviation of $\hat{\rho}_3^{\max}$), mode of the estimated maximal direction $(\hat{\alpha}_1, \hat{\alpha}_2, \hat{\alpha}_3)$ and $\hat{p}_\alpha$, the proportion of times in which the estimated direction was the mode, for 1000 simulated samples of size $n = 500, 1000, 5000$.

| C | $n$ | $\bar{\hat{\rho}}_{12}$ | $\bar{\hat{\rho}}_{13}$ | $\bar{\hat{\rho}}_{23}$ | $\bar{\hat{\rho}}_3^-$ | $\bar{\hat{\rho}}_3^+$ | $\bar{\hat{\rho}}_3^*$ | $\bar{\hat{\rho}}_3^{\max}$ | $\hat{\sigma}_{\rho_3^{\max}}$ | $\hat{\alpha}_1$ | $\hat{\alpha}_2$ | $\hat{\alpha}_3$ | $\hat{p}_\alpha$ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 500 | 0.778 | 0.359 | 0.358 | 0.530 | 0.466 | 0.498 | 0.530 | 0.028 | −1 | −1 | −1 | 1.00 |
| | 1000 | 0.777 | 0.362 | 0.361 | 0.531 | 0.469 | 0.500 | 0.531 | 0.019 | −1 | −1 | −1 | 1.00 |
| | 5000 | 0.778 | 0.361 | 0.361 | 0.532 | 0.469 | 0.500 | 0.532 | 0.009 | −1 | −1 | −1 | 1.00 |
| 2 | 500 | 0.942 | 0.483 | 0.679 | 0.723 | 0.680 | 0.701 | 0.723 | 0.023 | −1 | −1 | −1 | 1.00 |
| | 1000 | 0.942 | 0.481 | 0.677 | 0.722 | 0.679 | 0.700 | 0.722 | 0.016 | −1 | −1 | −1 | 1.00 |
| | 5000 | 0.943 | 0.483 | 0.679 | 0.723 | 0.680 | 0.702 | 0.723 | 0.007 | −1 | −1 | −1 | 1.00 |
| 3 | 500 | 0.848 | 0.460 | −0.010 | 0.422 | 0.443 | 0.433 | 0.443 | 0.027 | 1 | 1 | 1 | 0.93 |
| | 1000 | 0.848 | 0.459 | −0.012 | 0.421 | 0.442 | 0.432 | 0.442 | 0.019 | 1 | 1 | 1 | 0.98 |
| | 5000 | 0.849 | 0.460 | −0.011 | 0.422 | 0.443 | 0.433 | 0.443 | 0.009 | 1 | 1 | 1 | 1.00 |
| 4 | 500 | −0.759 | −0.242 | −0.315 | −0.439 | −0.438 | −0.439 | 0.288 | 0.021 | 1 | −1 | 1 | 0.83 |
| | 1000 | −0.760 | −0.242 | −0.316 | −0.439 | −0.439 | −0.439 | 0.283 | 0.017 | 1 | −1 | 1 | 0.92 |
| | 5000 | −0.761 | −0.241 | −0.316 | −0.439 | −0.439 | −0.439 | 0.280 | 0.008 | 1 | −1 | 1 | 1.00 |
| 5 | 500 | −0.259 | −0.452 | −0.452 | −0.358 | −0.417 | −0.387 | 0.245 | 0.027 | 1 | 1 | −1 | 1.00 |
| | 1000 | −0.258 | −0.453 | −0.452 | −0.358 | −0.417 | −0.388 | 0.245 | 0.019 | 1 | 1 | −1 | 1.00 |
| | 5000 | −0.259 | −0.453 | −0.453 | −0.358 | −0.418 | −0.388 | 0.245 | 0.009 | 1 | 1 | −1 | 1.00 |
| 6 | 500 | −0.109 | −0.073 | −0.783 | −0.311 | −0.332 | −0.322 | 0.296 | 0.022 | 1 | −1 | 1 | 0.66 |
| | 1000 | −0.109 | −0.074 | −0.783 | −0.311 | −0.333 | −0.322 | 0.290 | 0.015 | 1 | −1 | 1 | 0.73 |
| | 5000 | −0.109 | −0.073 | −0.784 | −0.312 | −0.333 | −0.322 | 0.284 | 0.008 | 1 | −1 | 1 | 0.93 |
| 7 | 500 | −0.147 | −0.668 | −0.075 | −0.283 | −0.310 | −0.297 | 0.265 | 0.024 | −1 | 1 | 1 | 0.84 |
| | 1000 | −0.144 | −0.668 | −0.077 | −0.283 | −0.309 | −0.296 | 0.260 | 0.019 | −1 | 1 | 1 | 0.99 |
| | 5000 | −0.146 | −0.668 | −0.077 | −0.284 | −0.310 | −0.297 | 0.259 | 0.009 | −1 | 1 | 1 | 1.00 |
| 8 | 500 | 0.985 | −0.746 | −0.839 | −0.200 | −0.200 | −0.200 | 0.861 | 0.014 | −1 | −1 | 1 | 1.00 |
| | 1000 | 0.985 | −0.748 | −0.840 | −0.201 | −0.201 | −0.201 | 0.861 | 0.010 | −1 | −1 | 1 | 1.00 |
| | 5000 | 0.985 | −0.747 | −0.840 | −0.201 | −0.200 | −0.200 | 0.859 | 0.004 | −1 | −1 | 1 | 1.00 |
| 9 | 500 | −0.760 | 0.430 | −0.854 | −0.394 | −0.395 | −0.395 | 0.686 | 0.019 | 1 | −1 | 1 | 1.00 |
| | 1000 | −0.760 | 0.429 | −0.854 | −0.395 | −0.395 | −0.395 | 0.685 | 0.015 | 1 | −1 | 1 | 1.00 |
| | 5000 | −0.761 | 0.430 | −0.854 | −0.395 | −0.395 | −0.395 | 0.683 | 0.006 | 1 | −1 | 1 | 1.00 |
| 10 | 500 | 0.261 | −0.451 | 0.451 | 0.058 | 0.116 | 0.087 | 0.243 | 0.028 | −1 | 1 | 1 | 1.00 |
| | 1000 | 0.258 | −0.452 | 0.453 | 0.057 | 0.116 | 0.086 | 0.245 | 0.020 | −1 | 1 | 1 | 1.00 |
| | 5000 | 0.259 | −0.452 | 0.453 | 0.057 | 0.116 | 0.087 | 0.245 | 0.009 | −1 | 1 | 1 | 1.00 |
| 11 | 500 | 0.072 | −0.110 | 0.784 | 0.238 | 0.259 | 0.248 | 0.297 | 0.020 | −1 | 1 | 1 | 0.67 |
| | 1000 | 0.074 | −0.109 | 0.783 | 0.239 | 0.260 | 0.249 | 0.290 | 0.015 | −1 | 1 | 1 | 0.73 |
| | 5000 | 0.074 | −0.109 | 0.784 | 0.239 | 0.260 | 0.250 | 0.284 | 0.008 | −1 | 1 | 1 | 0.91 |
| 12 | 500 | 0.146 | −0.667 | 0.076 | −0.161 | −0.135 | −0.148 | 0.265 | 0.023 | 1 | 1 | −1 | 0.82 |
| | 1000 | 0.147 | −0.667 | 0.075 | −0.162 | −0.135 | −0.148 | 0.261 | 0.018 | 1 | 1 | −1 | 0.89 |
| | 5000 | 0.146 | −0.668 | 0.077 | −0.162 | −0.135 | −0.148 | 0.259 | 0.009 | 1 | 1 | −1 | 1.00 |

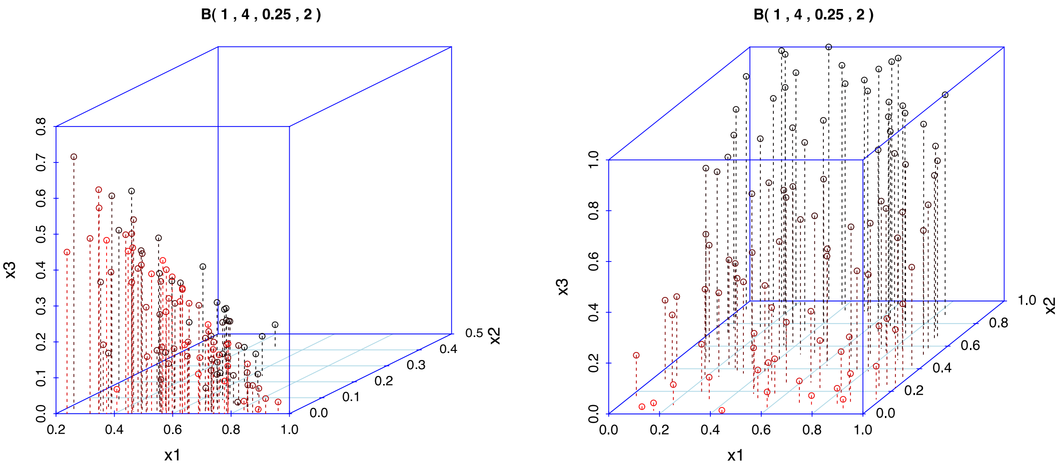


**Fig. 1.** 100 simulated samples for trivariate Beta distribution. Scatterplot for the simulated data with (left) original margins and (right) margins transformed to [0, 1] by scaling ranks.

Cases 10, 11 and 12 illustrate observation 4, with negative $\bar{\hat{\rho}}_{13}$. In the first two cases the 3-dimensional versions of Spearman's rho, $\bar{\hat{\rho}}_3^*$, $\bar{\hat{\rho}}_3^+$, $\bar{\hat{\rho}}_3^-$ are all positive. In the last, the 3-dimensional versions of Spearman's rho, $\bar{\hat{\rho}}_3^*$, $\bar{\hat{\rho}}_3^+$, $\bar{\hat{\rho}}_3^-$ are all negative. Fig. 2 (cases 11 and 12) shows that it may be hard to identify the direction of maximal dependence from

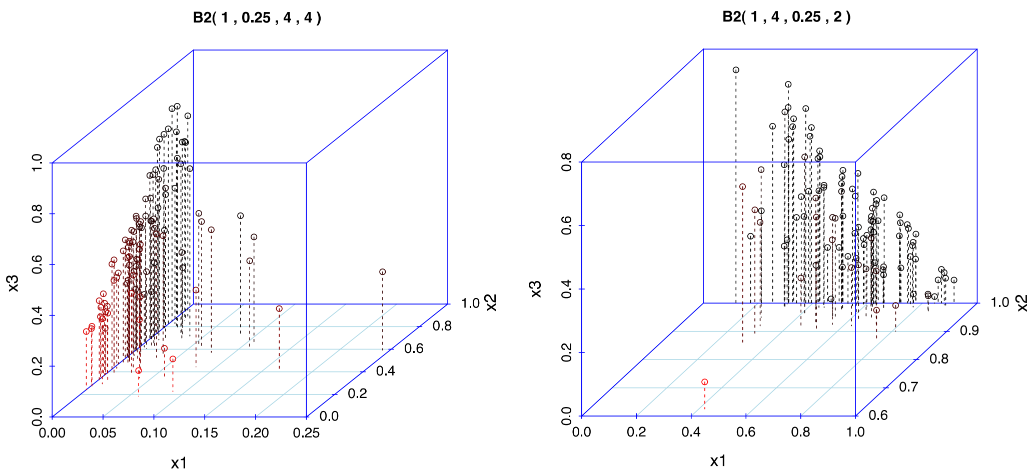


**Fig. 2.** 100 simulated samples for trivariate B2 distribution. Scatterplot for the simulated data with original margins.

the scatterplot of the simulated observations, when the data shows moderate/strong pairwise correlations, on the left $\bar{\hat{\rho}}_{23} = 0.784$ ($n = 5000$) and on the right $\bar{\hat{\rho}}_{13} = -0.668$ ($n = 5000$).

We observe that cases 4 through 9 and 12 illustrate that both $\bar{\hat{\rho}}_3^{+}$ and $\bar{\hat{\rho}}_3^{-}$ can be negative. Cases 10 and 11 illustrate that even when $\bar{\hat{\rho}}_3^{+}$ and $\bar{\hat{\rho}}_3^{-}$ are both positive $\bar{\hat{\rho}}_3^{\max}$ may be larger than either $\bar{\hat{\rho}}_3^{+}$ or $\bar{\hat{\rho}}_3^{-}$.

From Table 3, we see that the relationship between $\hat{\sigma}_{\rho_3^{\max}}$ and the sample size $n$ follows the $\frac{1}{\sqrt{n}}$ rule as expected from Theorem 5.3.

### 6.2. Application to a real data set

Our data consists of trivariate energy measures for 132 recorded sentences in English (EN), 216 sentences in French (FR) and 216 sentences in Catalan (CA), digitalized at 16.000 samples a second (i.e. sample rate of 16 kHz). This data comes from a corpus belonging to the *Laboratorie de Sciences Cognitives et Psycholinguistique (EHESS/CNRS)*. For each sentence, the three energy measurements correspond to the energy between 80 and 800 Hz, 820 and 1480 Hz and between 1500 and 5000 Hz respectively.

#### 6.2.1. Energy bands

Denote by $\vartheta_t^l(f)$ the power spectral density at time $t$ and frequency $f$, for language $l$, which is the square of the coefficient for frequency $f$ of the local Fourier decomposition of the speech signal. The time is discretized in steps of 2 ms and the frequency is discretized in steps of 20 Hz. The values of the power spectral density are estimated using a 25 ms Gaussian window.

The sentences $j, j = 1, \ldots, J^l$ ($J^l = 132$ if $l$ =EN, $J^l = 216$ if $l$ =FR or CA) are isolated phrases (not a running text) to guarantee the independence between them. For each sentence $j$ of length $T_j^l, j = 1, \ldots, J^l$, we consider the following stochastic processes, named energies, $t = 1, \ldots, T_j^l$,

$$\eta_1^{j,l}(t) = \sum_{f=80,100,\ldots,800} \vartheta_t^{j,l}(f), \qquad \eta_2^{j,l}(t) = \sum_{f=820,1520,\ldots,1480} \vartheta_t^{j,l}(f),$$
$$\eta_3^{j,l}(t) = \sum_{f=1500,1520,\ldots,5000} \vartheta_t^{j,l}(f).$$

Our measurements are the mean value energies along the sentence for each sentence $j$ of length $T_j^l$. That is, the random variables we will analyze are $E_1^l$, $E_2^l$ and $E_3^l$, where for each sentence $j$,

$$E_1^{j,l} = \frac{1}{T_j^l} \sum_{t=1,\ldots,T_j^l} \eta_1^{j,l}(t), \qquad E_2^{j,l} = \frac{1}{T_j^l} \sum_{t=1,\ldots,T_j^l} \eta_2^{j,l}(t), \qquad E_3^{j,l} = \frac{1}{T_j^l} \sum_{t=1,\ldots,T_j^l} \eta_3^{j,l}(t).$$

Fixed $l$, we assume that $(E_1^{j,l}, E_2^{j,l}, E_3^{j,l})$ are identically distributed for $j = 1, \ldots, J^l$. The frequencies for the bands were chosen based on previous works about automatic segmentations in vowels and consonants of the speech signal by Garcia et al. [14].

Abercrombie [2] claims that the languages are clustered into rhythmic classes, commanded by different rhythmic units, (a) syllable-timed class characterized by the syllabic intervals (supposed to be equal); (b) stress-timed class in which the unit is defined by the stress and (c) mora-timed class where the rhythmic unit is given by the mora, which is a sub-unit of

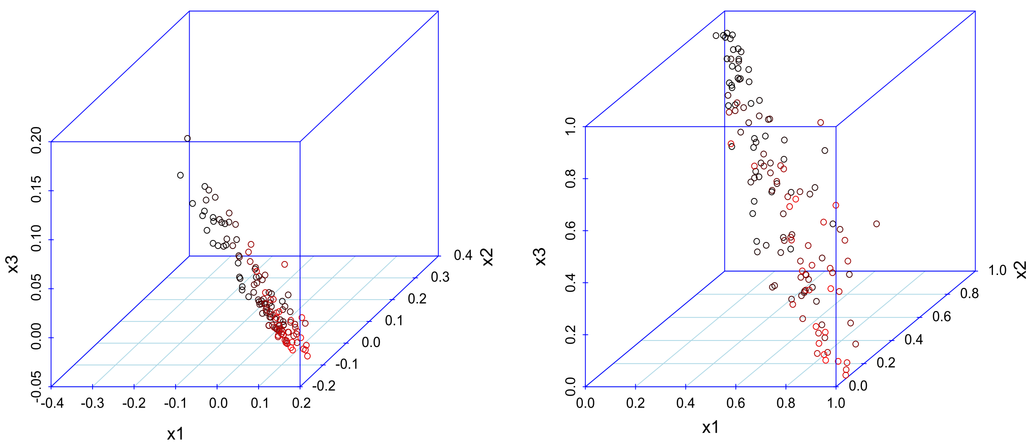


**Fig. 3.** Scatterplot for the trivariate energy data with (left) original margins and (right) margins transformed to [0, 1] by scaling ranks, for English.

**Table 4**
Estimated parameters for the trivariate energy data $(E_1, E_2, E_3)$, for English (EN), French (FR) and Catalan (CA).

| $l$ | $\hat{\rho}_{12}$ | $\hat{\rho}_{13}$ | $\hat{\rho}_{23}$ | $\hat{\rho}_3^-$ | $\hat{\rho}_3^+$ | $\hat{\rho}_3^*$ | $\hat{\rho}_3^{\max}$ | $(\alpha_1, \alpha_2, \alpha_3)$ |
|---|---|---|---|---|---|---|---|---|
| EN | −0.924 | −0.635 | 0.434 | −0.379 | −0.371 | −0.375 | 0.668 | $(1, -1, -1)$ |
| FR | −0.881 | −0.825 | 0.684 | −0.356 | −0.325 | −0.341 | 0.812 | $(1, -1, -1)$ |
| CA | −0.514 | −0.713 | 0.036 | −0.389 | −0.405 | −0.397 | 0.429 | $(-1, 1, 1)$ |

**Table 5**
Estimated parameters for the trivariate energy data $(-E_1^l, E_2^l, E_3^l)$ and $(E_1^l, -E_2^l, -E_3^l)$, for English.

| | $\hat{\rho}_{12}$ | $\hat{\rho}_{13}$ | $\hat{\rho}_{23}$ | $\hat{\rho}_3^-$ | $\hat{\rho}_3^+$ | $\hat{\rho}_3^*$ | $\hat{\rho}_3^{\max}$ | $(\alpha_1, \alpha_2, \alpha_3)$ |
|---|---|---|---|---|---|---|---|---|
| $(-E_1^l, E_2^l, E_3^l)$ | 0.924 | 0.635 | 0.434 | 0.668 | 0.660 | 0.664 | 0.668 | $(-1, -1, -1)$ |
| $(E_1^l, -E_2^l, -E_3^l)$ | 0.924 | 0.635 | 0.434 | 0.660 | 0.668 | 0.664 | 0.668 | $(1, 1, 1)$ |

the syllable. For example in Japanese, syllables with short vowels have one mora and syllables with long vowels have two or more morae. Dauer [3] as well as Ramus et al. [22] extract two main phonetic/phonologic properties and differences related to (a) and (b), the characteristics are (i) syllable structure: stress-timed languages have a greater variety of syllable types than syllable-timed languages and (ii) vowel reduction: in stress-timed languages, unstressed syllables usually have a reduced vocalic system. According to that, French and English are members of different classes, (a) and (b) respectively, while Catalan has a syllabic system according to a typically syllabic language but it has vowel reduction, i.e. Catalan mixes (i) and (ii) (see Ramus et al. [22]). Several correlates have been proposed for detecting historical changes in some language for example in Portuguese, see Frota et al. [8], for detecting differences between branches of Portuguese, see Galves et al. [13] and for detecting the existence of rhythmic classes, see for example Ramus et al. [22] and Garcia et al. [14]. Here we want to introduce a correlate based on the three band of energies (from the spectrogram). In specific, we aim to introduce as a correlate our 3-dimensional index of dependence which shows a new perspective to measure and understand the differences between the languages in function of bands of energies. We note that by conception (correlations of ranks of the observations) this index is resistant to natural differences in the quality/conditions of recording of each sentence and for each language. We conjectured that in general, there exists a compensation between the bands of energies for each language. More specifically, large values of $E_1^l$, tend to occur with small values of $E_2^l$ and $E_3^l$, because the majority of the phonemes show high values in the inferior band of energy.

### 6.2.2. Results

First of all we focus on English, to analyze in detail the results for this language. For English, the maximal directional coefficient is $\hat{\rho}_3^{\max} = 0.668$ in direction $(1, -1, -1)$ so that $\hat{\rho}_3^- = 0.668$ for the random variables $(-E_1^l, E_2^l, E_3^l)$ and $\hat{\rho}_3^+ = 0.668$ for the random variables $(E_1^l, -E_2^l, -E_3^l)$. This dependence is clearly visible in Fig. 3. We see the scatterplot for the random variables $(-E_1^l, E_2^l, E_3^l)$ in Fig. 4 (left) and for the random variables $(E_1^l, -E_2^l, -E_3^l)$ in Fig. 4 (right) and the estimated parameters in Table 5.

For all the languages the index is given by the equation $\hat{\rho}_3^{(1,-1,-1)} = \frac{2}{3}\hat{\rho}_{23} - \hat{\rho}_3^-$ or $\hat{\rho}_3^{(-1,1,1)} = \frac{2}{3}\hat{\rho}_{23} - \hat{\rho}_3^+$; in either case, we observe the relevance of the pairwise correlation $\hat{\rho}_{23}$ (the correlation between energy bands 2 and 3). In this way the index of maximal dependence is given by a transformation of that pairwise correlation and some contribution of $\hat{\rho}_3^-$ $(\hat{\rho}_3^+)$ depending on the language. From Table 4 we can verify that positive dependence is detected by $\rho_3^{\max}$ in the

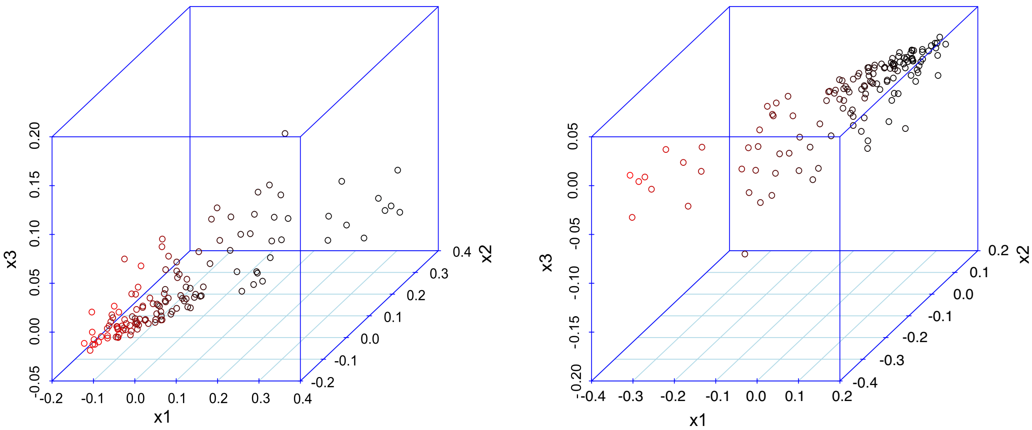


**Fig. 4.** Scatterplot for $(-E_1^l, E_2^l, E_3^l)$ on the left and $(E_1^l, -E_2^l, -E_3^l)$ on the right, for English.

direction $(1, -1, -1)$ in the case of English and French. This can be interpreted as "large" values of $E_1^l$ tend to occur with "small" values of $E_2^l$ and $E_3^l$, $l =$ EN, FR. However, the maximal positive dependence in the case of Catalan is verified in the direction $(-1, 1, 1)$, i.e. "small" values of $E_1^l$ tend to occur with "large" values of $E_2^l$ and $E_3^l$, $l =$ CA. The different ways that languages distribute the energy into the three bands could be used to improve the study of languages through their energies (see Garcia et al. [14]). It is advantageous to have a single calculation ($\rho_3^{\max}$) with remarkable statistical properties, rather than an ad-hoc procedure where one must perform 8 calculations to find the maximal correlation and its direction. In addition the new index identifies the direction of maximal dependence, and through it, we can track the composition of the index, pointing the contribution (in terms of magnitude of the index) of each one of the three pairwise Spearman's rho and 3-dimensional versions of Spearman's rho.

We use the bootstrap (see Schmid and Schmidt [23]) to estimate the standard deviation of $\hat{\rho}_3^{\max}$, and using a sample size equal to 500 was obtained $\hat{\sigma}_{\rho_3^{\max}} = 0.01$ for the 3 languages. In addition we computed the success rate of $(\alpha_1, \alpha_2, \alpha_3)$ (Table 4) that was 0.686, 0.926 and 0.79 for English, French and Catalan respectively. For the magnitude of $\hat{\rho}_3^{\max}$, we observe that for French, $\hat{\rho}_3^{\max}$ achieves the largest value followed by English while $\hat{\rho}_3^{\max}$ of Catalan achieves the least value among the 3 languages. We conjecture that syllable-timed languages can reach the highest values of $\rho_3^{\max}$, while the stress-timed languages can reach the lowest values. Mixed languages can achieve lower values, depending on the occurrence of the vowel reduction.

## 7. Conclusion

The index $\rho_3^{\max}$ of maximal dependence introduced in this paper to detect dependence in trivariate distributions has a simple expression as a function of the pairwise Spearman's rho coefficients and the three common 3-dimensional versions of Spearman's rho. The definition of $\rho_3^{\max}$ is based on the coefficients of directional dependence (see Nelsen and Úbeda-Flores [20]). Although $\rho_3^{\max}$ has nice properties such as normalization, invariance under permutations and monotone transformations, and continuity, it fails to be a measure of multivariate concordance. The existence of well-known estimators for the usual pairwise Spearman's rho coefficients and the three common 3-dimensional versions of Spearman's rho allows us to define similar estimators of $\rho_3^{\max}$ and the coefficients of directional dependence. We show in this paper that there exists an empirical process related to our index (similarly for the coefficients of directional dependence), that allows us to establish desirable properties for the estimator of the index, that is, it is asymptotically normal distributed, asymptotically unbiased and asymptotically consistent. Our simulation study exhibits cases where the direction of maximal dependence can be either easy or difficult to recognize by examining scatter-plots after replacing the data by ranks. The index $\rho_3^{\max}$ identifies positive dependence undetected by the existing 3-dimensional versions of Spearman's rho, for example, in cases where at least two of the pairwise Spearman's rho correlations are negative. We exhibit this situation in our simulation study and in a real data set.

The study of $\rho_3^{\max}$ has revealed some preliminary results that are beyond the scope of this paper. For example, Theorem 5.1 is true for an arbitrary dimension $d \geq 3$, as are Theorems 5.2 and 5.3. However, the geometric interpretations of the index $\rho_d^{\max}$, Theorem 3.1 and Table 1 need to be reformulated in dimensions higher than 3. To analyze $\rho_d^{\max}$ for $d > 3$ it is necessary to first investigate directional coefficients $\rho_d^{\alpha}$, generalizations of the coefficients $\rho_3^{\alpha}$ introduced in Nelsen and Úbeda-Flores [20]. Extending the results of this paper to construct indexes in higher dimensions based both on generalizations of Spearman's rho and other measures of association is the subject of future work.

## Acknowledgments

The authors gratefully acknowledge the financial support for this research provided by FAPESP grant 10/51940-5 to R. Nelsen, they also gratefully acknowledge the support from CNPq's projects 485999/2007-2 and 476501/2009-1 and USP project ''Mathematics, computation, language and the brain'' to J. García and V. González-López. The authors wish to thank three referees and an associate editor for their many helpful comments and suggestions on an earlier draft of this paper.

## References

[1] K. Aas, C. Czado, A. Frigessi, H. Bakken, Pair-copula constructions of multiple dependence, Insurance: Mathematics and Economics 44 (2) (2009) 182–198.
[2] D. Abercrombie, Elements of General Phonetics, Aldine, Chicago, 1967 (Chapter 5).
[3] R.M. Dauer, Stress-timing and syllable-timing reanalyzed, Journal of Phonetics 11 (1983) 51–62.
[4] P. Deheuvels, An asymptotic decomposition for multivariate distribution-free tests of independence, Journal of Multivariate Analysis 11 (1981) 102–113.
[5] A. Dolati, M. Úbeda-Flores, On measures of multivariate concordance, Journal of Probability and Statistical Sciences 4 (2006) 147–163.
[6] D. Dugué, Sur des tests d'indépendance indépendants de la loi, C. R. Acad. Sci. Paris Sér. A-B 281 (Aii) (1975) A1103–A1104.
[7] J.D. Fermanian, D. Radulovic, M. Wegkamp, Weak convergence of empirical copula processes, Bernoulli 10 (5) (2004) 847–860.
[8] S. Frota, C. Galves, M. Vigário, V.A. González-López, B. Abaurre, The phonology of rhythm from Classical to Modern Portuguese, Journal of Historical Linguistics 2 (2) (2012) 173–207.
[9] P. Gänßler, Empirical Processes. Hayward, CA: IMS Lecture Notes Monograph Series, Vol. 3, 1983.
[10] P. Gänßler, W. Stute, Seminar on Empirical Processes, DMV- Seminar, vol. 9, Birkhäuser Verlag, ISBN: 3-7643-1921-6, 1987.
[11] S. Gaißer, M. Ruppert, F. Schmid, A multivariate version of Hoeffding's Phi-Square, Journal of Multivariate Analysis 101 (10) (2010) 2571–2586.
[12] S. Gaißer, F. Schmid, On testing equality of pairwise rank correlations in a multivariate random vector, Journal of Multivariate Analysis 101 (10) (2010) 2598–2615.
[13] A. Galves, C. Galves, J. Garcia, N.L. Garcia, F. Leonardi, Context tree selection and linguistic rhythm retrieval from written texts, Annals of Applied Statistics 6 (1) (2012) 186–209.
[14] J. Garcia, U. Gut, A. Galves, Vocale — A Semi-Automatic Annotation Tool for Prosodic Research. Paper presented at Speech Prosody 2002, Aix-en-Provence (can be downloaded from http://aune.lpl.univ-aix.fr/sp2002/pdf/garcia-gut-galves.pdf) (date last viewed 11/19/11), 2002.
[15] C. Genest, J.F. Quessy, B. Rémillard, Asymptotic local efficiency of Cramér-Von Mises tests for Multivariate Independence, The Annals of Statistics 35 (1) (2007) 166–191.
[16] L.L.R. Rifo, V.A. González-López, Full Bayesian analysis for a model of tail dependence, Communications in Statistics. Theory and Methods 41 (22) (2012) 4107–4123.
[17] H. Joe, Multivariate concordance, Journal of Multivariate Analysis 35 (1990) 12–30.
[18] N.L. Johnson, S. Kotz, Distributions in Statistics: Continuous Multivariate Distributions, Wiley, New Yok, 1972.
[19] R.B. Nelsen, Nonparametric measures of multivariate association, in: L. Ruschendorf, B. Schweizer, M.D. Taylor (Eds.), Distributions with Given Marginals and Related Topics, vol. 28, IMS Lecture Notes-Monograph Series, Hayward, CA, 1996, pp. 223–232.
[20] R.B. Nelsen, M. Úbeda-Flores, Directional Dependence in Multivariate Distributions, Annals of the Institute of Statistical Mathematics 64 (2012) 677–685.
[21] J.F. Quessy, Theoretical efficiency comparisons of independence tests based on multivariate versions of Spearman's rho, Metrika 70 (2009) 315–338.
[22] F. Ramus, M. Nespor, J. Mehler, Correlates of linguistic rhythm in the speech signal, Cognition 73 (3) (1999) 265–292.
[23] F. Schmid, R. Schmidt, Bootstrapping Spearman's multivariate rho, COMPSTAT, Proceedings in Computational Statistics (2006) 759–766.
[24] F. Schmid, R. Schmidt, Multivariate extensions of Spearman's rho and related statistics, Statistics and Probability Letters 77 (2007) 407–416.
[25] J. Segers, Asymptotics of empirical copula processes under nonrestrictive smoothness assumptions, Bernoulli 18 (3) (2012) 764–782.
[26] M.D. Taylor, Multivariate measures of concordance, Annals of the Institute of Statistical Mathematics 59 (2007) 789–806.
[27] M.D. Taylor, Some properties of multivariate measures of concordance, arXiv:0808.3105 [math.PR], 2008.
[28] A.W. Van der Vaart, J.A. Wellner, Weak Convergence and Empirical Processes, Springer, New York, 1996.
[29] D. Wied, H. Dehling, M. Van Kampen, D. Vogel, A fluctuation test for constant Spearman's rho with nuisance-free limit distribution (Preprint), 2011.