A CLASS OF WEIGHTED RANK CORRELATION MEASURES

BY

MAJID SANATGAR (YAZD), ALI DOLATI (YAZD), AND
MOHAMMAD AMINI (MASHHAD)

Abstract. We propose a class of weighted rank correlation measures extending Spearman's rho. This class consists of two types of measures. The first type, which extends Blest's rank correlation, places more emphasis on the agreement in top ranks. The second one places more emphasis on the agreement in the bottom ranks. The asymptotic distribution of the proposed measures and some of their properties are studied. A simulation study is performed to compare the performance of the proposed statistics for testing independence by using asymptotic relative efficiency calculations.

2020 Mathematics Subject Classification: Primary 62H20; Secondary 62H05.

Key words and phrases: weighted rank correlation, copula, asymptotic distribution.

1. INTRODUCTION

There are certain cases in which n objects are ranked by two independent sources, where the interest is focused on agreement in the top rankings and disagreements in items at the bottom of the rankings, or vice versa. For example, every year a large number of students apply for higher education. The graduate committee of the university may like to choose the best candidates based on some criteria such as GPA (grade point averange) and the average of their grades in the major courses that they passed during their bachelor's level. In such cases, to minimize the cost of interviewing all of the candidates, a measure that gives more weights to those with higher grades is required. Measures of rank correlation, such as Spearman's rho and Kendall's tau, are not quite appropriate for calculating the correlation between two sets of ranks, since all ranks are handled equally. This problem leads to the definition of the *Weighted Rank Correlation* (WRC) measures, which give greater priority to the agreement in top or bottom ranking. Salama and Quade [16], [14] first studied a WRC of two sets of rankings, sensitive to agreements in the top rankings and ignored disagreements in the other items in a certain degree. For an

application in sensitivity analysis, Iman and Conover [7] proposed the top-down concordance coefficient, which centers on the agreement in the top rankings. Shieh [18] studied a weighted version of Kendall's tau that could reflect well the importance of agreement in the top ranking. Blest [2] introduced a rank correlation measure that gives more weights to the top rankings. Pinto da Costa and Soares [13] proposed a WRC measure that weights the distance between two ranks by using a linear function of those ranks, giving more importance to higher ranks than lower ones. Maturi and Abdelfattah [10] proposed a WRC measure with different weights to emphasize the agreement of the top rankings. Coolen-Maturi [3] extended this index to more than two sets of rankings, but again the focus was only on the agreement in the top or bottom rankings. The behavior of several WRC measures derived from Spearman's rank correlation was investigated by Dancelli, Manisera, and Vezzoli [4].

By starting from the formula for Spearman's rank correlation measure, this paper proposes a general class of WRC measures that weight the distance between two sets of ranks. Two types of weights, which are polynomial functions of the ranks, are considered to construct some new WRC measures. The first type, which extends Blest's rank correlation, places more emphasis on the agreement in the top rankings. The second one places more emphasis on the bottom ranks.

The rest of the paper is organized as follows. The proposed WRC measures are introduced in Section 2. The weighted correlation coefficients, which are the population versions of the proposed WRC measures, are introduced in Section 3. The quantiles of the proposed WRC measures for small samples and their asymptotic distributions for large samples are presented in Section 4. In Section 5, a simulation study is performed to compare the performance of the proposed statistics for testing independence by using the asymptotic relative efficiency and the empirical powers of the tests. Finally, some discussions and possible extensions are given in Section 6.

2. THE PROPOSED WRC MEASURES

Let $(X_1, Y_1), \ldots, (X_n, Y_n)$ be a random sample of size n from a continuous bivariate distribution and let $(R_1, S_1), \ldots, (R_n, S_n)$ denote the corresponding vectors of ranks. Spearman's well-known rank correlation measure is given by

$$(2.1) \qquad \rho_{n,s} = \frac{12}{n^3 - n} \sum_{i=1}^{n} R_i S_i - \frac{3(n+1)}{n-1}.$$

The drawback of Spearman's rank correlation is that, in some cases, it measures the correlation between two sets of ranks without highlighting an important part of the ranks. For example, consider three consumers A, B, and C who ranked nine aspects of a product by attributing "1" to the most important aspect and "9" to the least important one. Their rankings are given in Table 1. As we see, the top ranks of

TABLE 1. Rankings of a product by three consumers

A	1	2	3	4	5	6	7	8	9
B	**1**	**2**	**3**	9	8	7	6	4	5
C	5	6	4	3	2	1	**7**	**8**	**9**

(A, B) are more similar than those of (A, C), and the bottom ranks of (A, C) are more similar than those of (A, B), but Spearman's rank correlation gives the same value 0.433 for two sets of rankings (A, B) and (A, C). For the cases where the differences in the top ranks would seem to be more critical, Blest [2] suggested that these discrepancies should be emphasized. He proposed an alternative measure of rank correlation that attaches more significance to the early ranking of an initially given order. Blest's index is defined by

$$(2.2) \qquad \gamma_n = \frac{2n+1}{n-1} - \frac{12}{n^2-1} \sum_{i=1}^{n} \left(1 - \frac{R_i}{n+1}\right)^2 S_i.$$

The values of Blest's index for two sets of rankings (A, B) and (A, C) are 0.597 and 0.27, respectively. As we see, in situations such as the rankings (A, B) where the top ranks should be emphasized, Blest's index reflects the emphasis on the agreement of top rankings, and it is therefore an appropriate rank correlation measure.

In the following, we develop a general theory for weighted rank correlation measures by giving suitable weights to the distance between two sets of ranks. Let $D_i = S_i - R_i$, $i = 1, \ldots, n$. The most common form of Spearman's rank correlation coefficient between two sets of rankings $R_1, \ldots, R_n$ and $S_1, \ldots, S_n$ is given by (see [8])

$$(2.3) \qquad \rho_{n,s} = 1 - \frac{2\sum_{i=1}^{n} D_i^2}{\max(\sum_{i=1}^{n} D_i^2)},$$

where $\max(\sum_{i=1}^{n} D_i^2) = (n^3 - n)/3$ represents the value of the summation when there is a perfect discordance between rankings, that is, $S_i = n + 1 - R_i$, $i = 1, \ldots, n$. Throughout the rest of the paper, we assume, without loss of generality, that the sample pairs are given in accordance with the increasing magnitude of X components, so that $R_i = i$ for $i = 1, \ldots, n$ and $D_i = S_i - i$. According to Blest's idea [2], if the set of points $(0, 0)$, $(\sum_{i=1}^{k}(n+1-i), \sum_{i=1}^{k}(S_i - i))$, $k = 1, \ldots, n$, is determined in the coordinate plane, then Spearman's rho is obtained by adding the width of the given points, and normalization, as follows:

$$(2.4) \qquad \rho_{n,s} = 1 - \frac{2\sum_{k=1}^{n}\sum_{i=1}^{k}(S_i - i)}{\max(\sum_{k=1}^{n}\sum_{i=1}^{k}(S_i - i))}.$$

By changing the order of summation, it is easy to see that $\sum_{k=1}^{n}\sum_{i=1}^{k}(S_i - i) = \frac{1}{2}\sum_{i=1}^{n} D_i^2$. While (2.4) and (2.3) are two different representations of ordinary

Spearman's rho, Blest's index is the normalized version of the area that appears from connecting the points to each other. By looking again at Blest's index, one can imagine that the bars $\eta_k = \sum_{i=1}^{k}(S_i - i), \ k = 1, \ldots, n$, are assigned certain weights, in comparison to Spearman's rho that does not give any weight to that bar. Now we consider a general class of WRC measures of the form

$$(2.5) \qquad \nu_n = 1 - \frac{2\sum_{i=1}^{n} w_i \eta_i}{\max(\sum_{i=1}^{n} w_i \eta_i)},$$

where the positive constants w_i are suitable weights. To construct WRC measures that are sensitive to agreement in top rankings (lower ranks), for $p = 1, 2, \ldots$ and $n > 1$, we choose the weights $w_i = (n + 1 - i)^p - (n - i)^p$. The class of WRC measures provided by (2.5) is then

$$(2.6) \qquad \nu_{n,p}^{(l)} = 1 + \frac{2\sum_{i=1}^{n}(i - S_i)(n + 1 - i)^p}{\sum_{i=1}^{n}(n + 1 - 2i)(n + 1 - i)^p}.$$

Alternatively, by choosing the weights $w_i = i^p - (i - 1)^p$, one can obtain measures that are sensitive to agreement in bottom rankings (upper ranks). In this case, the class of WRC measures (2.5) is simplified to

$$(2.7) \qquad \nu_{n,p}^{(u)} = 1 + \frac{2\sum_{i=1}^{n}(i - S_i)(n^p - (i - 1)^p)}{\sum_{i=1}^{n}(n + 1 - 2i)(n^p - (i - 1)^p)}.$$

Let $\kappa_{n,p} = \sum_{i=1}^{n} i^p$. It is easily seen that

$$\sum_{i=1}^{n}(n + 1 - 2i)(n + 1 - i)^p = 2\kappa_{n,p+1} - (n + 1)\kappa_{n,p},$$

$$\sum_{i=1}^{n}(n + 1 - 2i)(n^p - (i - 1)^p) = 2\kappa_{n-1,p+1} - (n - 1)\kappa_{n-1,p}.$$

The coefficients (2.6) and (2.7) could be rewritten in terms of $\kappa_{n,p}$ as

$$(2.8) \qquad \nu_{n,p}^{(l)} = \frac{(n + 1)\kappa_{n,p} - 2\sum_{i=1}^{n} S_i(n + 1 - i)^p}{2\kappa_{n,p+1} - (n + 1)\kappa_{n,p}},$$

$$(2.9) \qquad \nu_{n,p}^{(u)} = \frac{-(n + 1)\kappa_{n-1,p} + 2\sum_{i=1}^{n} S_i(i - 1)^p}{2\kappa_{n-1,p+1} - (n - 1)\kappa_{n-1,p}}.$$

Note that for $p = 1$, both $\nu_{n,p}^{(l)}$ and $\nu_{n,p}^{(u)}$ reduce to Spearman's rank correlation coefficient (2.1). For $p = 2$, the measure $\nu_{n,p}^{(l)}$ reduces to Blest's rank correlation coefficient (2.2). The coefficients $\nu_{n,p}^{(l)}$ and $\nu_{n,p}^{(u)}$ are asymmetric WRC measures,

that is, the correlation of (X, Y) is not the same as that of (Y, X). One can obtain the symmetrized version of (2.6) as follows:

$$
\begin{aligned}
\nu_{n,p}^{(s.l)} &= \frac{\nu_{n,p}^{(l)}(X, Y) + \nu_{n,p}^{(l)}(Y, X)}{2} \\
&= \frac{(n+1)\kappa_{n,p} - \sum_{i=1}^{n}[S_i(n+1-i)^p + i(n+1-S_i)^p]}{2\kappa_{n,p+1} - (n+1)\kappa_{n,p}}.
\end{aligned}
$$

Similarly the symmetrized version of (2.7) is given by

$$
\nu_{n,p}^{(s.u)} = \frac{-(n+1)\kappa_{n-1,p} + \sum_{i=1}^{n}[S_i(i-1)^p + i(S_i-1)^p]}{2\kappa_{n-1,p+1} - (n-1)\kappa_{n-1,p}}.
$$

For $p = 1$, the WRC measures $\nu_{n,p}^{(s.l)}$ and $\nu_{n,p}^{(s.u)}$ are equal to Spearman's rank correlation (2.1). For $p = 2$, the measure $\nu_{n,p}^{(s.l)}$ is the symmetrized version of Blest's index (2.2). Table 2 shows the values of $\nu_{n,p}^{(l)}$ and $\nu_{n,p}^{(u)}$, $p = 1, 2, 3, 4, 5$, and their symmetrized versions for the rankings (A, B) and (A, C) given in Table 1. The result

TABLE 2. Values of Spearman's rho and WRC measures for three sets of rankings in Table 1

	(A, C)				(A, B)			
p	$\nu_{n,p}^{(l)}$	$\nu_{n,p}^{(u)}$	$\nu_{n,p}^{(s.l)}$	$\nu_{n,p}^{(s.u)}$	$\nu_{n,p}^{(l)}$	$\nu_{n,p}^{(u)}$	$\nu_{n,p}^{(s.l)}$	$\nu_{n,p}^{(s.u)}$
1	0.433	0.433	0.433	0.433	0.433	0.433	0.433	0.433
2	0.270	0.637	0.263	0.645	0.596	0.229	0.603	0.220
3	0.155	0.768	0.140	0.776	0.720	0.112	0.728	0.094
4	0.081	0.851	0.057	0.858	0.808	0.045	0.815	0.016
5	0.033	**0.905**	0.001	**0.910**	**0.869**	0.006	**0.875**	0.032

illustrates the sensitivity of these indices to the agreement in top and bottom rankings. We note that $\nu_{n,p}^{(l)}$, $\nu_{n,p}^{(u)}$, and their symmetrized versions take values in $[-1, 1]$. In particular, the value of these measures is 1 when $S_i = i$ (a perfect positive dependency between two sets of ranks) and it is -1 when $S_i = n + 1 - i$ (a perfect negative dependency).

3. WEIGHTED CORRELATION COEFFICIENTS

In this section, we introduce weighted correlation coefficients $\nu_p^{(l)}$ and $\nu_p^{(u)}$ and their symmetrized versions $\nu_p^{(s.l)}$ and $\nu_p^{(s.u)}$ as the population counterparts of the WRC measures $\nu_{n,p}^{(l)}$, $\nu_{n,p}^{(u)}$, $\nu_{n,p}^{(s.l)}$, and $\nu_{n,p}^{(s.u)}$. Each of these coefficients can be expressed as a linear functional of the quantity $A(u, v) = C(u, v) - \Pi(u, v)$, where C is the copula [20] associated with the pair (X, Y) and $\Pi(u, v) = uv$ is the copula

of independent random variables. For $p = 2, 3, \ldots$, we have

$$\nu_p^{(l)} = 2(p+1)(p+2) \int_0^1 \int_0^1 (1-u)^{p-1}(C(u,v) - uv)\, du\, dv,$$

$$\nu_p^{(u)} = 2(p+1)(p+2) \int_0^1 \int_0^1 u^{p-1}(C(u,v) - uv)\, du\, dv,$$

(3.1)
$$\nu_p^{(s.u)} = (p+1)(p+2) \int_0^1 \int_0^1 (u^{p-1} + v^{p-1})(C(u,v) - uv)\, du\, dv,$$

$$\nu_p^{(s.l)} = (p+1)(p+2)$$
$$\times \int_0^1 \int_0^1 ((1-u)^{p-1} + (1-v)^{p-1})(C(u,v) - uv)\, du\, dv.$$

Note that for $p = 1$, all of these coefficients reduce to Spearman's rho given by

$$\rho_s = 12 \int_0^1 \int_0^1 (C(u,v) - uv)\, du\, dv.$$

For $p = 2$, the coefficient $\nu_p^{(l)}$ reduces to Blest's correlation coefficient [5] given by

$$\gamma = 24 \int_0^1 \int_0^1 (1-u)C(u,v)\, du\, dv - 2.$$

REMARK 3.1. A probabilistic interpretation can be given for the weighted correlation coefficients $\nu_p^{(l)}$ and $\nu_p^{(u)}$ and their symmetrized versions $\nu_p^{(s.l)}$ and $\nu_p^{(s.u)}$. We provide an illustration for $\nu_p^{(l)}$. For $p = 1, 2, \ldots$, consider the cumulative distribution function

$$F_p(u,v) = (1 - (1-u)^p)v, \quad 0 \leqslant u, v \leqslant 1.$$

Let (U, V) be a random vector with joint distribution function F_p. For a copula C, the coefficient $\nu_p^{(l)}$ has the following representation:

$$\nu_p^{(l)} = \frac{2(p+1)(p+2)}{p} \int_0^1 \int_0^1 (C(u,v) - uv)\, dF_p(u,v)$$
$$= \frac{E_{F_p}[C(U,V) - \Pi(U,V)]}{E_{F_p}[M(U,V) - \Pi(U,V)]},$$

where $M(u,v) = \min(u,v)$ and E_{F_P} denotes the expectation with respect to F_p. Thus, the coefficient $\nu_p^{(l)}$ can be considered as an average distance between the copula C and the independent copula Π, where the average is taken with respect to the bivariate distribution function F_p. The proposed weighted correlation coefficients could be seen as average quadrant dependent (AQD) measures of association studied in [1].

For $\nu_p^{(l)}$, it is more convenient to use the following alternative representation:

$$(3.2) \qquad \nu_p^{(l)} = \frac{2(p+1)(p+2)}{p} \int\limits_0^1 \int\limits_0^1 (1-u)^p(1-v)\,dC(u,v) - \frac{p+2}{p},$$

which follows from the fact that

$$\int\limits_0^1 \int\limits_0^1 C(u,v)\,dF_p(u,v) = P(W \leqslant U, Z \leqslant V)$$
$$= P(U \geqslant W, V \geqslant Z)$$
$$= \int\limits_0^1 \int\limits_0^1 \bar{F}_p(u,v)\,dC(u,v),$$

where (W, Z) and (U, V) are two independent pairs distributed as the copula C and the joint distribution F_p, respectively, and $\bar{F}_p(u,v) = P(U > u, V > v) = (1-u)^p(1-v)$ is the survival function associated with F_p. An alternative representation for $\nu_p^{(u)}$ is given by

$$(3.3) \qquad \nu_p^{(u)} = \frac{2(p+1)(p+2)}{p} \int\limits_0^1 \int\limits_0^1 (1-u^p)(1-v)dC(u,v) - (p+2).$$

In the following examples, we provide the values of the weighted correlation coefficients $\nu_p^{(l)}$ and $\nu_p^{(u)}$ for some copulas.

EXAMPLE 3.1. Let C_θ be a member of the Cuadras–Augé family of copulas [11] given by

$$(3.4) \qquad C_\theta(u,v) = [\min(u,v)]^\theta [uv]^{1-\theta}, \quad \theta \in [0,1].$$

This family of copulas is positively ordered in $\theta \in [0,1]$, that is, for $\theta_1 \leqslant \theta_2$, we have $C_{\theta_1}(u,v) \leqslant C_{\theta_2}(u,v)$ for all $u, v \in [0,1]$. This family has no lower tail dependence, whereas the upper tail dependence parameter is given by $\lambda_U = \theta$; see [11]. For this family of copulas, we have

$$(3.5) \qquad \nu_p^{(u)} = \frac{(p+2)(p^2+2p-3+\theta)}{p(p+3-\theta)},$$

$$(3.6) \qquad \nu_p^{(l)} = \frac{\theta(p+2)\left(1-p(p+1)B(p,4-\theta)\right)}{p(2-\theta)},$$

where $B(a,b) = \int_0^1 \int_0^1 x^{a-1}(1-x)^{b-1}\,dx$ is the beta function. For every $\theta \in [0,1]$, $\nu_p^{(u)}$ is increasing in p and $\nu_p^{(l)}$ is decreasing in p. In particular, for this family

of copulas, for every $\theta \in [0, 1]$ and $p = 2, 3, \ldots$,

$$\nu_p^{(l)} \leqslant \nu_1^{(l)} = \rho_s = \nu_1^{(u)} \leqslant \nu_p^{(u)}.$$

EXAMPLE 3.2. Let C_θ be a member of the Raftery family of copulas [11] given by

$$C_\theta(u, v) = \min(u, v) + \frac{1 - \theta}{1 + \theta} (uv)^{\frac{1}{1-\theta}} \left\{ 1 - [\max(u, v)]^{-\frac{1+\theta}{1-\theta}} \right\}, \quad \theta \in [0, 1].$$

This family is also positively ordered in $\theta \in [0, 1]$ and has no upper tail dependence, whereas the lower tail dependence parameter is given by $\lambda_L = \frac{2\theta}{\theta+1}$; see [11]. It is easy to see that for this family of copulas, for every $\theta \in [0, 1]$ and $p = 2, 3, \ldots$, we have

$$\nu_p^{(u)} \leqslant \nu_1^{(u)} = \rho_s = \nu_1^{(l)} \leqslant \nu_p^{(l)}.$$

4. THE QUANTILES AND ASYMPTOTIC DISTRIBUTION OF WRC MEASURES

In general, the asymptotic behavior of the proposed WRC measures can be studied by applying some standard results from the theory of empirical processes [21]. First, we mention an asymptotic formula for $\kappa_{n,p} = \sum_{i=1}^{n} i^p$, which we need in what follows. By the definition of the Riemann integral, we have

$$(4.1) \qquad \frac{\kappa_{n,p}}{n^{p+1}} = \frac{1}{n} \sum_{i=1}^{n} \left(\frac{i}{n} \right)^p = \int_0^1 x^p \, dx + O(n^{-1}).$$

Let $(X_1, Y_1), \ldots, (X_n, Y_n)$ be a random sample of size n from a pair (X, Y) of continuous random variables with joint distribution function H, marginal distribution functions F and G, and the associated copula C. Let $(1, S_1), (2, S_2), \ldots, (n, S_n)$ be the ranks of the rearranged sample. It is known (see Rüschendorf [15]) that the copula C can be estimated by the empirical copula defined for all $u, v \in [0, 1]$ by

$$C_n(u, v) = \frac{1}{n} \sum_{i=1}^{n} I\left(\frac{i}{n + 1} \leqslant u, \frac{S_i}{n + 1} \leqslant v \right),$$

where $I(A)$ denotes the indicator function of the set A. The empirical versions of the weighted correlation coefficients $\nu_p^{(l)}$ and $\nu_p^{(u)}$ and their symmetrized versions $\nu_p^{(s.l)}$ and $\nu_p^{(s.u)}$ defined by (3.1) could be written in terms of C_n. By plugging C_n in (3.2), the empirical version of $\nu_p^{(l)}$ is of the form

$$\tilde{\nu}_{n,p}^{(l)} = \frac{2(p+1)(p+2)}{p} \int_0^1 \int_0^1 (1 - u)^p (1 - v) \, dC_n(u, v) - \frac{p+2}{p}.$$

By using the representation (2.8) and the identity (4.1), straightforward calculations give

$$(4.2) \quad \tilde{\nu}_{n,p}^{(l)} = \frac{2(p+1)(p+2)}{np} \sum_{i=1}^{n} \left(1 - \frac{i}{n+1}\right)^p \left(1 - \frac{S_i}{n+1}\right) - \frac{p+2}{p}$$

$$= \frac{2(p+1)(p+2)\kappa_{n,p}}{np(n+1)^p} - \frac{2(p+1)(p+2)}{np(n+1)^{p+1}} \sum_{i=1}^{n} S_i(n+1-i)^p - \frac{p+2}{p}$$

$$= \frac{(p+1)(p+2)\kappa_{n,p}}{np(n+1)^p} + \frac{(p+1)(p+2)(2\kappa_{n,p+1} - (n+1)\kappa_{n,p})}{np(n+1)^{p+1}} \nu_{n,p}^{(l)}$$

$$- \frac{p+2}{p}$$

$$= (1 + O(n^{-1}))\nu_{n,p}^{(l)} + O(n^{-1}).$$

By using (3.3), a similar argument shows that the empirical version of the coefficient $\nu_p^{(u)}$ is given by

$$(4.3) \quad \tilde{\nu}_{n,p}^{(u)} = \frac{2(p+1)(p+2)}{np} \sum_{i=1}^{n} \left(1 - \left(\frac{i}{n+1}\right)^p\right) \left(1 - \frac{S_i}{n+1}\right) - (p+2)$$

$$= (1 + O(n^{-1}))\nu_{n,p}^{(u)} + O(n^{-1}).$$

In the following, we provide the asymptotic distribution of the WRC measures $\nu_{n,p}^{(l)}$, $\nu_{n,p}^{(u)}$ and their symmetrized versions $\nu_{n,p}^{(s.l)}$, $\nu_{n,p}^{(s.u)}$. As shown by Segers [17], C_n converges weakly to C as $n \to \infty$ whenever C is regular, that is, when the partial derivatives $C_1(u,v) = \partial C(u,v)/\partial u$ and $C_2(u,v) = \partial C(u,v)/\partial v$ exist everywhere on $[0,1]^2$ and C_1 and C_2 are continuous on $(0,1) \times [0,1]$ and $[0,1] \times (0,1)$, respectively. Moreover, the empirical copula process $\mathbb{C}_n = \sqrt{n}(C_n - C)$ converges weakly as $n \to \infty$ to a centered Gaussian process $\hat{\mathbb{C}}$ on $[0,1]^2$, defined for all $u,v \in [0,1]$ by

$$(4.4) \quad \hat{\mathbb{C}}(u,v) = \mathbb{C}(u,v) - \frac{\partial}{\partial u}C(u,v)\mathbb{C}(u,1) - \frac{\partial}{\partial v}C(u,v)\mathbb{C}(1,v),$$

where $\mathbb{C}(u,v)$ is the Brownian bridge on $[0,1]^2$ with covariance function

$$E(\mathbb{C}(u,v)\mathbb{C}(s,t)) = C(\min(u,s),\min(v,t)) - C(u,v)C(s,t).$$

THEOREM 4.1. *Suppose that C is a regular copula. Then $\sqrt{n}\,(\nu_{n,p}^{(l)} - \nu_p^{(l)})$, $\sqrt{n}\,(\nu_{n,p}^{(u)} - \nu_p^{(u)})$, $\sqrt{n}\,(\nu_{n,p}^{(s.u)} - \nu_p^{(s.u)})$, and $\sqrt{n}\,(\nu_{n,p}^{(s.l)} - \nu_p^{(s.l)})$ are asymptotically centered normal with asymptotic variances given by*

$$(\sigma_p^{(l)})^2 = 4(p+1)^2(p+2)^2$$
$$\times \int\limits_{[0,1]^4} (1-u)^{p-1}(1-s)^{p-1}E(\hat{\mathbb{C}}(u,v)\hat{\mathbb{C}}(s,t))\,du\,dv\,ds\,dt,$$

$$(\sigma_p^{(u)})^2 = 4(p+1)^2(p+2)^2$$
$$\times \int\limits_{[0,1]^4} u^{p-1}s^{p-1}E(\hat{\mathbb{C}}(u,v)\hat{\mathbb{C}}(s,t))\,du\,dv\,ds\,dt,$$

$$(4.5) \quad (\sigma_p^{(s.u)})^2 = (p+1)^2(p+2)^2$$
$$\times \int\limits_{[0,1]^4} (u^{p-1}+v^{p-1})(s^{p-1}+t^{p-1})E(\hat{\mathbb{C}}(u,v)\hat{\mathbb{C}}(s,t))\,du\,dv\,ds\,dt,$$

$$(\sigma_p^{(s.l)})^2 = (p+1)^2(p+2)^2$$
$$\times \int\limits_{[0,1]^4} ((1-u)^{p-1}+(1-v)^{p-1})((1-s)^{p-1}+(1-t)^{p-1})$$
$$\times E(\hat{\mathbb{C}}(u,v)\hat{\mathbb{C}}(s,t))\,du\,dv\,ds\,dt,$$

where $\hat{\mathbb{C}}$ is the Gaussian process defined by (4.4).

Proof. We prove the result for $\sqrt{n}\,(\nu_{n,p}^{(l)} - \nu_p^{(l)})$; similar arguments hold for the other quantities. From (4.2), we have

$$\sqrt{n}\,(\nu_{n,p}^{(l)} - \nu_p^{(l)}) = (1 + O(n^{-1}))\sqrt{n}\,(\tilde{\nu}_{n,p}^{(l)} - \nu_p^{(l)}) + O(n^{-1/2})$$

$$= 2(1+O(n^{-1}))(p+1)(p+2)\int\limits_0^1\int\limits_0^1 (1-u)^{p-1}[\sqrt{n}\,(C_n(u,v)-C(u,v))]\,du\,dv$$

$$+ O(n^{-1/2}).$$

Since the integral on the right side is a continuous linear functional of the empirical copula process, the left hand side is asymptotically centered normal with asymptotic variance $(\sigma_p^{(l)})^2$, as stated in the theorem. ∎

COROLLARY 4.1. *Consider the null hypothesis of independence, that is, $C(u,v) = uv$. Then $\sqrt{n}\,\nu_{n,p}^{(l)}$, $\sqrt{n}\,\nu_{n,p}^{(u)}$, $\sqrt{n}\,\nu_{n,p}^{(s.l)}$, and $\sqrt{n}\,\nu_{n,p}^{(s.u)}$ are asymptotically centered normal with the asymptotic standard deviations*

$$\sigma_p^{(l)} = \sigma_p^{(u)} = \frac{(p+2)}{\sqrt{3(2p+1)}} \quad and \quad \sigma_p^{(s.l)} = \sigma_p^{(s.u)} = \sqrt{\frac{p^2 + 10p + 7}{6(2p+1)}}.$$

Proof. For $C(u,v) = uv$, the covariance function of the limiting Gaussian process $\hat{\mathbb{C}}$ takes the form $E(\hat{\mathbb{C}}(u,v)\hat{\mathbb{C}}(s,t)) = (\min(u,s) - us)(\min(v,t) - vt)$. An application of Theorem 4.1 and a routine integration give the result. ∎

In order to use the proposed WRC measures for testing independence, one needs to find their distributions or the quantiles of the distribution under the hypothesis of independence. The following result provides the expectation and variance of $\nu_{n,p}^{(l)}$. A similar result could be found for $\nu_{n,p}^{(u)}$, $\nu_{n,p}^{(s.l)}$, and $\nu_{n,p}^{(s.u)}$.

THEOREM 4.2. *Under the hypothesis of independence between two sets of ranks, it follows that*

$$E(\nu_{n,p}^{(l)}) = 0, \quad \mathrm{var}(\nu_{n,p}^{(l)}) = \frac{n(n+1)}{3} \, \frac{\kappa_{n,2p} - \frac{1}{n}(\kappa_{n,p})^2}{(2\kappa_{n,p+1} - (n+1)\kappa_{n,p})^2}.$$

Proof. We note from (2.6) that the WRC measure $\nu_{n,p}^{(l)}$ can be written as a linear combination of the linear rank statistics of the form $a_n + b_n \sum_{i=1}^{n} a(i, S_i)$, where

$$a_n = 1 + \left(2 \sum_{i=1}^{n} i(n+1-i)^p \right) \left(\sum_{i=1}^{n} (n+1-2i)(n+1-i)^p \right)^{-1},$$

$$b_n = -2 \left(\sum_{i=1}^{n} (n+1-2i)(n+1-i)^p \right)^{-1},$$

and $a(i, S_i) = S_i(n+1-i)^p$. The mean and the variance of the quantity $S = \sum_{i=1}^{n} a(i, S_i)$ can be obtained, for example, by using [19, Theorem 1, p. 57]. See also [6]. ∎

5. COMPARING THE PERFORMANCE OF INDEPENDENCE TESTS

In this section, we compare the Pitman asymptotic relative efficiency (or Pitman ARE) and the empirical power of tests of independence based on the proposed WRC measures.

5.1. Pitman efficiency. Consider a parametric family $\{C_\theta\}$ of copulas with $\theta = \theta_0$ corresponding to the independence case. Let T_{1n} and T_{2n} be two test statistics for testing $H_0 : \theta = \theta_0$ versus $H_1 : \theta > \theta_0$ that reject the null hypothesis for large values of T_{1n} and T_{2n}. Suppose that T_{1n} and T_{2n} satisfy the regularity conditions:

(1) There exist continuous functions $\mu_i(\theta)$ and $\sigma_i(\theta)$, $\theta > \theta_0, i = 1, 2$, such that for all sequences $\theta_n = \theta_0 + h/\sqrt{n}$ and $h > 0$,

$$\lim_{n \to \infty} P_{\theta_n} \left(\frac{\sqrt{n} \, (T_{in} - \mu_i(\theta_n))}{\sigma_i(\theta_n)} < z \right) = \Phi(z), \quad z \in \mathbb{R}, \, i = 1, 2,$$

where $\Phi(z)$ is the standard normal distribution function.

(2) The function $\mu_i(\theta)$ is continuously differentiable at $\theta = \theta_0$ and $\mu_i'(\theta_0) > 0$, $\sigma_i(\theta_0) > 0, i = 1, 2$.

Under these conditions, the Pitman ARE of T_{n1} relative to T_{n2} is equal to

$$\text{ARE}(T_{1n}, T_{2n}) = \left[\frac{\frac{\partial}{\partial\theta}\mu_1\big|_{\theta=\theta_0}}{\frac{\partial}{\partial\theta}\mu_2\big|_{\theta=\theta_0}} \cdot \frac{\sigma_2(\theta_0)}{\sigma_1(\theta_0)}\right]^2;$$

for more details, see [12]. In the following, we compare the ARE of the proposed WRC measures relative to Spearman's rho for the Cuadras–Augé family of copulas given by (3.4).

EXAMPLE 5.1. Suppose that the copula of (X, Y) is a member of the Cuadras–Augé family of copulas given by (3.4). Let $T_{1n} = \nu_{n,p}^{(l)}$ and $T_{2n} = \rho_{ns}$. By Theorem 4.1, for the test statistics based on the WRC measure $\nu_{n,p}^{(l)}$, the regularity conditions (1) and (2) are satisfied with $\theta_0 = 0, \mu_p(\theta) = \nu_p^{(l)}$, and σ_p^l given in Corollary 4.1 (for $\theta_0 = 0$). By using (3.6) and differentiation with respect to θ, one gets

$$\frac{d\nu_p^{(l)}}{d\theta}(0) = \frac{p+5}{2(p+3)}.$$

Since $\nu_1^{(l)} = \rho_s$ and $\nu_{n,1}^{(l)} = \rho_{n,s}$, from Corollary 4.1 we have

$$\text{ARE}(\nu_{n,p}^{(l)}, \rho_{ns}) = \frac{4(p+5)^2(2p+1)}{3(p+2)^2(p+3)^2}.$$

Similarly, for $\nu_{n,p}^{(s.l)}$, $\nu_{n,p}^{(u)}$, and $\nu_{n,p}^{(s.u)}$, we have

$$\text{ARE}(\nu_{n,p}^{(s.l)}, \rho_{n,s}) = \frac{8(p+5)^2(2p+1)}{3(p+3)^2(p^2+10p+7)},$$
$$\text{ARE}(\nu_{n,p}^{(u)}, \rho_{n,s}) = \frac{16(2p+1)}{3(p+3)^2},$$
$$\text{ARE}(\nu_{n,p}^{(s.u)}, \rho_{n,s}) = \frac{32(p+2)^2(2p+1)}{3(p+3)^2(p^2+10p+7)}.$$

Figure 1 shows the ARE of the test of independence based on WRC measures compared to the test based on Spearman's ρ for the Clayton copula as a family of copulas with lower tail dependence, and the Cuadras–Augé family of copulas as a family of copulas with upper tail dependence. As we see, the measure $\nu_{n,11}^{(s.l)}$ for the Clayton family of copulas, and the measure $\nu_{n,3}^{(s.u)}$ for the Cuadras–Augé family of copulas, have the largest Pitman ARE. The same results are still true by using Kendall's τ instead of Spearman's ρ since $\text{ARE}(T_\tau, T_\rho) = 1$.
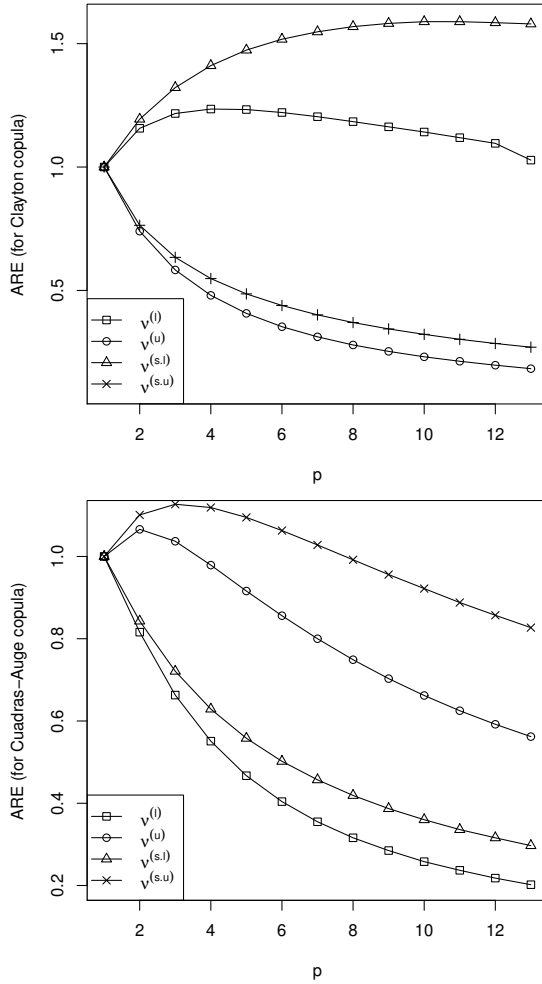
FIGURE 1. ARE of tests of independence based on WRC measures $\nu_{n,p}^{(l)}$, $\nu_{n,p}^{(u)}$, $\nu_{n,p}^{(s.l)}$, and $\nu_{n,p}^{(s.u)}$, for $p = 1, 2, \ldots, 13$, relative to Spearman's rho for the Clayton copula (top) and the Cuadras–Augé copula (bottom).

5.2. Power study. In the following, we compare the power of tests based on the WRC measures $\nu_{n,p}^{(s.l)}$ and $\nu_{n,p}^{(s.u)}$ with the tests based on Kendall's tau for testing independence against the positive quadrant dependence [9], that is,

$$H_0 : C(u,v) = uv \quad \text{against} \quad H_1 : C(u,v) > uv.$$

Monte Carlo simulations were carried out for the Gumbel and Clayton copulas for various degrees of dependence (in terms of Spearman's ρ) with a sample of size $n = 50$ at significance level 0.05. Figure 2 shows the empirical power of tests that are obtained under alternatives, defined by the Clayton and Gumbel copulas.
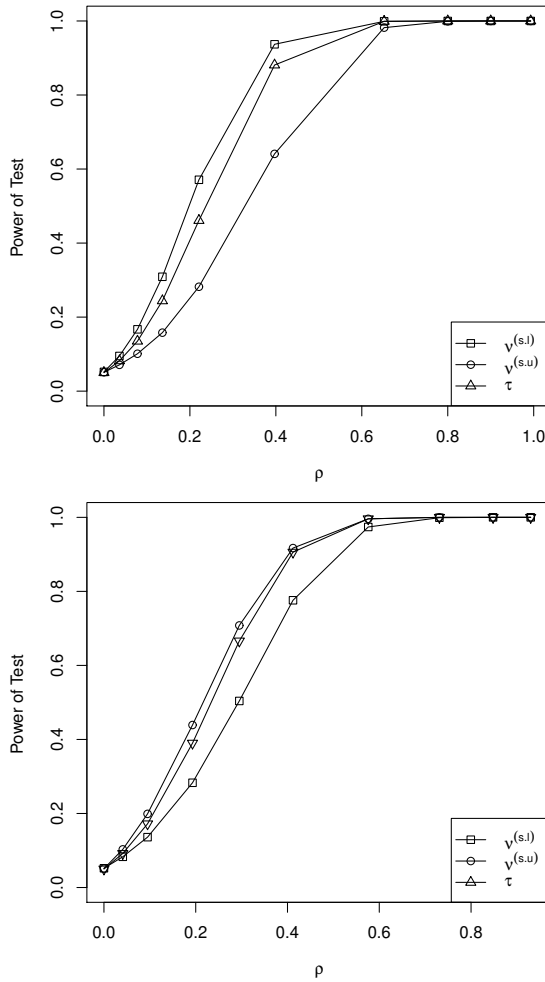
FIGURE 2. The power of tests of independence based on Kendall's tau (τ) and the WRC measures $\nu_5^{(s.l)}$ and $\nu_5^{(s.u)}$, computed from 50,000 samples of size 50 for the Clayton copula (bottom) and the Gumbel copula (top) with different levels of dependence in terms of Spearman's rho (ρ).

The Clayton copula has lower tail dependence and the Gumbel copula has upper tail dependence. For Clayton's family of copulas for all degrees of dependence the test based on $\nu_{n,5}^{(s.l)}$ has the maximum power. For the Gumbel family of copulas, the test based on $\nu_{n,5}^{(s.u)}$ has the maximum power. It seems that the members of the class $\nu_{n,p}^{(s.u)}$ perform very well compared with Kendall's tau (and Spearman's rho) if there exists a higher dependence in the upper tail. If there exists a higher dependence in the lower tail, the members of the class $\nu_{n,p}^{(s.l)}$ have a better performance. For a radially symmetric copula like the normal copula, the behavior of all tests of independence is the same.

6. DISCUSSION

In this paper, we presented a class of weighted rank correlation measures extending Spearman's rank correlation coefficient. The class was constructed by giving suitable weights to the distance between two sets of ranks to place more emphasis on items with low rankings than those with high rankings, or vice versa. The asymptotic distributions of the proposed measures in general and under the null hypothesis of independence are derived. We also carried out a simulation study to compare the performance of the test of independence based on the proposed measures with Kendall's rank correlation measure. Another line of research is the extension of the result to the situations where n objects are ranked by $m > 2$ independent sources, and our interest was focused on agreement in the bottom or top rankings.

REFERENCES

[1] J. Behboodian, A. Dolati and M. Úbeda-Flores, *Measures of association based on average quadrant dependence*, J. Probab. Statist. Sci. 3 (2005), 161–173.

[2] D. C. Blest, *Rank correlation an alternative measure*, Austral. New Zealand J. Statist. 42 (2000), 101–111.

[3] T. Coolen-Maturi, *A new weighted rank coefficient of concordance*, J. Appl. Statist. 41 (2014), 1721–1745.

[4] L. Dancelli, M. Manisera and M. Vezzoli, *On two classes of weighted rank correlation measures deriving from the Spearman's ρ*, in: Statistical Models for Data Analysis, Springer, (2013), 107–114.

[5] C. Genest and J. F. Plante, *On Blest's measure of rank correlation*, Canad. J. Statist. 31 (2003), 35–52.

[6] J. Hájek, *A Course in Nonparametric Statistics*, Holden-Day, San Francisco, 1969.

[7] R. L. Iman and W. J. Conover, *A measure of top-down correlation*, Technometrics 29 (1987), 351–357.

[8] M. Kendall, *Rank Correlation Methods*, Griffin, London, 1948.

[9] P. Matula, *A note on some inequalities for certain clases of positively dependent random variables*, Probab. Math. Statist. 24 (2004), 17–26

[10] T. A. Maturi and E. H. Abdelfattah, *A new weighted rank correlation*, J. Math. Statist. 4 (2008), 226–230.

[11] R. B. Nelsen, *An Introduction to Copulas*, 2nd ed., Springer, New York, 2006.

[12] Ya. Yu. Nikitin, *Asymptotic Efficiency of Nonparametric Tests*, Cambridge Univ. Press, 2009.

[13] J. Pinto da Costa and C. Soares, *A weighted rank measure of correlation*, Austral. New Zealand J. Statist. 47 (2005), 515–529.

[14] D. Quade and I. Salama, *A survey of weighted rank correlation*, in: Order Statistics and Nonparametrics: Theory and Applications, 1992, 213–224.

[15] L. Rüschendorf, *Asymptotic distributions of multivariate rank order statistics*, Ann. Statist. 4 (1976), 912–923.

[16] I. Salama and D. Quade, *A nonparametric comparison of two multiple regressions by means of a weighted measure of correlation*, Comm. Statist. Theory Methods 11 (1982), 1185–1195.

[17] J. Segers, *Asymptotics of empirical copula processes under non-restrictive smoothness assumptions*, Bernoulli 18 (2012), 764–782.

[18] G. S. Shieh, *A weighted Kendall's tau statistic*, Statist. Probab. Lett. 39 (1998), 17–24.

[19] Z. Šidák, P. K. Sen, and J. Hájek, *Theory of Rank Tests*, Academic Press, 1999.

[20] A. Sklar, *Fonctions de répartition à n dimensions et leurs marges*, Publ. Inst. Statist. Univ. Paris 8 (1959), 229–231.

[21] A. W. van der Vaart and J. A. Wellner, *Weak Convergence and Empirical Processes: with Applications to Statistics*, Springer, New York, 1996.

Majid Sanatgar   Ali Dolati
Department of Statistics   Department of Statistics
Yazd University   Yazd University
Yazd, Iran   Yazd, Iran
E-mail: m.sanatgar@stu.yazd.ac.ir   *E-mail*: adolati@yazd.ac.ir

Mohammad Amini
Department of Statistics
Ferdowsi University of Mashhad
Mashhad, Iran
E-mail: m-amini@um.ac.ir