On measuring and testing the ordinal correlation between valued outranking relations

Raymond Bisdorff
University of Luxembourg, FSTC/CSC/ILIAS
raymond.bisdorff@uni.lu

Abstract. We generalize Kendall's rank correlation measure τ to valued relations. Motivation for this work comes from the need to measure the level of approximation that is required when replacing a given valued outranking relation with a convenient crisp ordering recommendation.

Keywords: Multiple criteria decision aid; Ordinal correlation; Kendall's tau, Outranking relations; Bipolar credibility valuation.

Introduction

When proposing a measure for providing information on the potentially conflicting nature of the criteria in a given MCDA problem [1], we applied Kendall's rank correlation measure τ to the ordinal comparison of the marginal rankings observed on each criterion. Now, we propose to furthermore generalize the same idea to the direct comparison of bipolarly-valued binary relations [5, 6].

This work is motivated, first, by the need to finetune meta-heuristics for multiple criteria based clustering, where the eventual clustering results may be compared to an a priori given pairwise global outranking relation [2]. A second, similar motivation comes from the need to compare multiple criteria based rankings obtained with different ranking rules like Kemeny's, Kohler's, the PROMETHEE net flows rule [14], or, more recently, Dias-Lamboray's prudent leximin rule [3]. Assessing the operational performance of these rules may be based on the more or less consistent ordinal correlation observed between each ranking results and the empirical underlying valued outranking relation.

The present work is closely related to, without being inspired from, recent results concerning the formal and empirical analysis of the fuzzy gamma rank correlation coefficient [4].

After the formal introduction of our correlation measure, and the discussion of some of its properties, we provide empirical results for statistically testing the presence or absence of any correlation between different types of random relations, and more particularly, valued outrankings.

1 Measuring ordinal correlations

1.1 Ordinal correlation between crisp relations

Let R_1 and R_2 be two binary relations defined on the same finite set X of dimension n. Kendall's rank, or ordinal, correlation measure τ is essentially based on the idea of counting the number of concordant (equivalent) non reflexive pairwise relational situations, normalized by the total number $n(n-1)$ of possible such relational situations. If $C = \#\{\,(x,y) \in X^2 \;:\; x \neq y \text{ and } \big((x\,\mathrm{R}_1\,y) \Leftrightarrow (x\,\mathrm{R}_2\,y)\big)\,\}$ denotes the number of concordant non reflexive relational situations we observe, that Kendall's τ measure can be defined as follows[1]:

$$\tau(\mathrm{R}_1, \mathrm{R}_2) \quad := \quad 2 \times \frac{C}{n(n-1)} \;-\; 1\,. \qquad (1)$$

It is worthwhile noticing that Kendall [7, 8] used a very natural way (see [5, 6]) of transforming a direct counting of concordant, i.e. logically equivalent situations, into a bipolarly valued correlation index. Unanimously (100% equivalent situations) concordant relations are matched to a correlation index of value +1.0, 50% concordance between the relations (50% equivalent and 50% not equivalent situations) is matched to a zero-valued correlation index, and unanimously discordant relations

[1] Originally, Kendall [7, 8] counted the number of inversions observed when comparing two linear orders. Formula (1), hence, takes a dual form: $\tau(\mathrm{R}_1, \mathrm{R}_2) = 1 - 2 \times \big(n(n-1) - C\big) \,/\, n(n-1)$ [9, see page 104].

(100% non equivalent situations) are matched to a correlation index of value: -1.0.

Example 1.1. Let us consider the following crisp relations R_1 and R_2 defined on a set $X = \{a, b, c\}$ of nodes, where $R_1 = \{(b, c), (c, a)\}$ and $R_2 = \{(a, b), (a, c), (b, c)\}$. As we observe as many concordant pairs: (b, a), (b, c), and (c, b), as discordant pairs: (a, b), (a, c), and (c, a), the Kendall $\tau(\mathrm{R}_1, \mathrm{R}_2)$ correlation index equals: $\tau(\mathrm{R}_1, \mathrm{R}_2) = 2 \times \frac{3}{6} - 1 = 0.0$.

The τ rank correlation index implicitly relies on the assumption that each relation is completely determined. Either $(x \,\mathrm{R}\, y)$ or $\neg(x \,\mathrm{R}\, y)$; all relational situations between any pair of elements of X are exactly known. But, what happens when we compare now valued relations, where the validation of relational situations might be more or less precarious?

1.2 Valued equivalence of relational situations

Let R_1 and R_2 be two binary relations defined on the same finite set X of dimension n and characterized via a bipolar characteristic function r taking values in the rational interval $[-1.0; 1.0]$ [5, 6]. We call such relations, for short, *r-valued* and *of order n*.

For any such valued relation R, its characteristic function r supports the following semantics:

i) $r(x \,\mathrm{R}\, y) = \pm 1.0$ signifies that the relational situation $x \,\mathrm{R}\, y$ is certainly valid (+1.0), resp. invalid (-1.0);
ii) $r(x \,\mathrm{R}\, y) > 0.0$ signifies that the relational situation $x \,\mathrm{R}\, y$ is more valid than invalid;
iii) $r(x \,\mathrm{R}\, y) < 0.0$ signifies that the relational situation $x \,\mathrm{R}\, y$ is more invalid than valid;
iv) $r(x \,\mathrm{R}\, y) = 0.0$ signifies that the relational situation $x \,\mathrm{R}\, y$ is indeterminate, i.e. neither valid, nor invalid.

Logical negation, conjunction, and disjunction of such r-characteristic values may be respectively computed with changing the sign, applying a min, or max operator [5, 6, 10]. For instance:

$$\begin{aligned} r\big(\neg(x \,\mathrm{R}\, y)\big) &= -r(x \,\mathrm{R}\, y), \\ r\big((x \,\mathrm{R}_1\, y) \wedge (x \,\mathrm{R}_2\, y)\big) &= \min\big(r(x \,\mathrm{R}_1\, y), r(x \,\mathrm{R}_2\, y)\big), \\ r\big((x \,\mathrm{R}_1\, y) \vee (x \,\mathrm{R}_2\, y)\big) &= \max\big(r(x \,\mathrm{R}_1\, y), r(x \,\mathrm{R}_2\, y)\big). \end{aligned}$$

These logical operators, now, allow us to compute, for instance, the r-valued logical equivalence of any two relational situations:

$$\begin{aligned} &r\big(\ (x \,\mathrm{R}_1\, y) \Leftrightarrow (x \,\mathrm{R}_2\, y)\ \big) \\ &\quad = r\Big[\big(\neg(x \,\mathrm{R}_1\, y) \vee (x \,\mathrm{R}_2\, y)\big) \wedge \big(\neg(x \,\mathrm{R}_2\, y) \vee (x \,\mathrm{R}_1\, y)\big)\Big] \\ &\quad = \min\Big[\max\big(-r(x \,\mathrm{R}_1\, y), r(x \,\mathrm{R}_2\, y)\big), \\ &\qquad\qquad \max\big(r(-x \,\mathrm{R}_2\, y), r(x \,\mathrm{R}_1\, y)\big)\Big] \end{aligned}$$

Finally, we will need to measure the average level of determinateness of an r-valued relation R of order n, denoted $d(\mathrm{R})$, and taking value in the interval $[0; 1]$:

$$d(\mathrm{R}) := \frac{\sum_{x,y \in X^2}^{x \neq y} \mathrm{abs}(r(x \,\mathrm{R}\, y))}{n(n-1)}. \quad (2)$$

Thus, a *crisp* – a completely ± 1-valued – relation shows a determinateness degree of 1, whereas an *indeterminate* – a completely 0-valued – relation shows a determinateness degree of 0.

Example 1.2. We may apply the concepts and tools of this r-valued credibility calculus for assessing, for instance, the actual equivalence of the relational situations we observed in Example 1.1. Take for instance the situation $(a \,\mathrm{R}\, b)$. Here we have: $r(a \,\mathrm{R}_1\, b) = -1.0$ and $r(a \,\mathrm{R}_2\, b) = 1.0$. It follows that $r(a \,\mathrm{R}_1\, b \Leftrightarrow a \,\mathrm{R}_2\, b) = \min(-1.0, 1.0) = -1.0$. Whereas, if we take the pair (b, c), we obtain $r(b \,\mathrm{R}_1\, c \Leftrightarrow b \,\mathrm{R}_2\, c) = \min(1.0, 1.0) = 1.0$. Hence, we faithfully recover in the crisp case, the original Kendall τ values. Suppose now that relation R_1 is not certainly determined and $r(a \,\mathrm{R}_1\, b) = -\alpha$ with $\alpha \in [0; 1]$. In this case $r(a \,\mathrm{R}_1\, b \Leftrightarrow a \,\mathrm{R}_2\, b) = \min(-\alpha, 1.0) = -\alpha$. Similarly, suppose now that $r(b \,\mathrm{R}_1\, c) = \alpha$. In that case $r(b \,\mathrm{R}_1\, c \Leftrightarrow b \,\mathrm{R}_2\, c) = \min(\alpha, 1.0) = \alpha$.

This gives us a hint that the r-valued equivalence of two valued relational situations verifies the following important property:

Property 1.1.
Let R_1 and R_2 be any two r-valued relations defined on the same set X. For all x, y in X^2, we have:

$$\begin{aligned} &r\big(\ (x \,\mathrm{R}_1\, y) \Leftrightarrow (x \,\mathrm{R}_2\, y)\ \big) = \\ &\quad \pm \min\big(\ \mathrm{abs}(r(x \,\mathrm{R}_1\, y)), \mathrm{abs}(r(x \,\mathrm{R}_2\, y))\ \big). \end{aligned}$$

Proof.
Suppose $r(x \,\mathrm{R}_1\, y) = \alpha$ and $r(x \,\mathrm{R}_2\, y) = \beta$ with $\alpha, \beta \in [-1; 1]$. If $\mathrm{abs}(r(x \,\mathrm{R}_1\, y)) = \mathrm{abs}(r(x \,\mathrm{R}_2\, y))$, Property 1.1 follows immediately from Equation (2). Otherwise, we may observe the following cases:

1. $|\alpha| > |\beta|$:
   i) if $\alpha > \beta \geq 0$ then $\min[\max(-\alpha, \beta), \max(-\beta, \alpha)] = \beta > 0$;
   ii) if $\alpha > 0 > \beta$ then $\min[\max(-\alpha, \beta), \max(-\beta, \alpha)] = \beta < 0$;
   iii) if $\beta > 0 > \alpha$ then $\min[\max(-\alpha, \beta), \max(-\beta, \alpha)] = -\beta < 0$;
   iv) if $0 \leq \beta > \alpha$ then $\min[\max(-\alpha, \beta), \max(-\beta, \alpha)] = -\beta > 0$.
2. $|\beta| > |\alpha|$:
   i) if $\beta > \alpha \geq 0$ then $\min[\max(-\alpha, \beta), \max(-\beta, \alpha)] = \alpha > 0$;
   ii) if $\alpha > 0 > \beta$ then $\min[\max(-\alpha, \beta), \max(-\beta, \alpha)] = -\alpha < 0$;

iii) if $\beta > 0 > \alpha$ then
$\min[\max(-\alpha, \beta), \max(-\beta, \alpha)] = \alpha < 0$;

iv) if $0 \leq \alpha > \beta$ then
$\min[\max(-\alpha, \beta), \max(-\beta, \alpha)] = -\alpha > 0$.

□

With Property 1.1 in mind, we may now generalize Kendall's ordinal correlation measure for taking into account genuine r-valued relations.

1.3 Correlations between valued relations

The r-valued equivalence of relational situations may be judiciously used, as in the crisp case, for assessing the numerator of the ordinal correlation measure. Yet, stating the adequate denominator needs some further going considerations. In the classical crisp case, following Kendall, we divide the sum of pairwise equivalences with $n(n-1)$, i.e. the total number of concerned non reflexive situations. If we would proceed this way in the valued case, the resulting measure would integrate a mixture of both the ordinal correlation as well as the actual determinateness of the equivalence observed between the considered r-valued relations. To factor out both these effects we take, instead, as denominator the maximum possible sum of r-valued equivalences we could potentially observe when both r-valued relations would show completely concordant relational situations.

Hence, we formulate the r-valued ordinal correlation measure τ between two r-valued relations R_1 and R_2, defined on a same set X, as follows: $\tau(\mathrm{R}_1, \mathrm{R}_2) :=$

$$\frac{\sum_{x \neq y} r\big((x \, \mathrm{R}_1 \, y) \Leftrightarrow (x \, \mathrm{R}_2 \, y) \big)}{\sum_{x \neq y} \min \Big[\mathrm{abs}\big(r(x \, \mathrm{R}_1 \, y)\big), \mathrm{abs}\big(r(x \, \mathrm{R}_2 \, y)\big) \Big]} \tag{3}$$

where, in order to avoid divisions by zero, we assume that a zero sum of r-valued equivalences occurring in the numerator always takes strong precedence over the potential zero sum determinateness occurring in the denominator. Indeed, if the sum of absolute values of r-valued equivalences is zero, then so must essentially be the sum of the corresponding signed r-valued equivalences.

It is furthermore worthwhile noticing that the denominator in Formula (3), once divided by the number of non reflexive relational situations, i.e.:

$$\frac{\sum_{x \neq y} \min \Big[\mathrm{abs}\big(r(x \, \mathrm{R}_1 \, y)\big), \mathrm{abs}\big(r(x \, \mathrm{R}_2 \, y)\big) \Big]}{n(n-1)} \tag{4}$$

gives, in fact, the average determinateness degree of the r-valued equivalence relation $\mathrm{R}_1 \Leftrightarrow \mathrm{R}_2$ observed between both r-valued relations. In case of crisp relations, this determinateness degree always takes maximum value 1.0. But, as soon as one of both valued relations appears completely indeterminate, $d(\mathrm{R}_1 \Leftrightarrow \mathrm{R}_2)$ becomes 0. In this latter case, $\tau(\mathrm{R}_1, \mathrm{R}_2)$ becomes equally 0. Otherwise, $\tau(\mathrm{R}_1, \mathrm{R}_2)$ gives the ordinal correlation measure independently of their equivalence determinateness level.

Describing the ordinal correlation between two r-valued binary relations, hence, requires to show both, the relative ordinal correlation measure τ defined in Equation (3), as well as the determinateness degree D of the corresponding relational equivalence defined in Equation (2).

Example 1.3. To illustrate this insight, we consider in Table 1, two randomly r-valued relations R_1 and R_2 of order $n = 3$ and defined on a same set $X = \{a, b, c\}$. The pairwise r-valued equivalence situations $\mathrm{R}_1 \Leftrightarrow \mathrm{R}_2$ are shown in Table 2.

Table 1. Examples of randomly valued relations

$r(x \, \mathrm{R}_1 \, y)$	a	b	c
a	–	+0.68	+0.35
b	−0.94	–	+0.80
c	−1.00	+0.36	–
$r(x \, \mathrm{R}_2 \, y)$	a	b	c
a	–	−0.32	+0.58
b	−0.14	–	+0.75
c	−1.00	+0.08	–

Table 2. r-valued equivalence between R_1 and R_2

$r(x \, \mathrm{R}_1 \, y \Leftrightarrow x \, \mathrm{R}_2 \, y)$	a	b	c
a	–	−0.32	+0.35
b	+0.14	–	+0.75
c	+1.00	+0.08	–

Hence,

$$\begin{aligned} &\tau(\mathrm{R}_1, \mathrm{R}_2) \\ &= \frac{-0.32 + 0.35 + 0.14 + 0.75 + 1.00 + 0.08}{+0.32 + 0.35 + 0.14 + 0.75 + 1.00 + 0.08} \\ &= \frac{0.200}{0.264} = +0.7575\,, \end{aligned}$$

whereas the corresponding equivalence determinateness:

$$d(\mathrm{R}_1 \Leftrightarrow \mathrm{R}_2) = \frac{0.264}{6} = 0.44\,.$$

Thus, nearly 76% or the jointly determined ordinal information is actually shared by both r-valued relations, independently of the respective 44% of determinateness of the r-valued equivalence situations.

If we had instead used the classical denominator $n(n-1)$ for computing the actual correlation measure, we would have obtained a much smaller τ value of only: $\frac{0.200}{6} = +1/3$ (33.33% instead of 75.75%); potentially misleading us, thus, on the apparent correlation between R_1 and R_2. Notice that this result of $1/3$ is in fact the product of $\tau(\mathrm{R}_1, \mathrm{R}_2)$ with $d(\mathrm{R}_1, \mathrm{R}_2)$, i.e. 0.7575×0.44.

1.4 Properties of the ordinal correlation measure

Again, let R_1 and R_2 be two r-valued binary relations defined on a set X of dimension n. We say that R_1 and R_2 show *a same*, respectively *an opposite*, orientation if, for all non reflexive pairs (x, y) in X, $r(x\,\mathrm{R}_1\,y \;\Leftrightarrow\; x\,\mathrm{R}_2\,y) > 0$, respectively $r(x\,\mathrm{R}_1\,y \;\Leftrightarrow\; x\,\mathrm{R}_2\,y) < 0$.

Property 1.2. *If two r-valued relations R_1 and R_2, defined on the same set X, show a same, respectively an opposite, orientation, $\tau(\mathrm{R}_1, \mathrm{R}_2)$ equals $+1.0$, respectively -1.0, independently of their equivalence determinateness $d(\mathrm{R}_1, \mathrm{R}_2)$.*

Proof. Property 1.2 readily follows from Property 1.1 and the observation that in case of a same orientation, respectively an opposite orientation, the sum of terms in the numerator of Formula (3) equals the sum, respectively the negation, of the sum of terms in the denominator. □

The *logical negation* of an r-valued relation R, denoted $\neg\,\mathrm{R}$, is called its *dual* relation. And, the *reciprocal* of an r-valued relation R, denoted Я, is called its *converse* relation. The following very natural properties are verified by the generalized ordinal correlation measure τ.

Property 1.3. *Let R_1 and R_2 be two r-valued binary relations defined on a same set X and let the ordinal correlation measure τ be defined by Formula (3):*

$$\tau(\mathrm{R}_1, \mathrm{R}_2) = \tau(\mathrm{R}_2, \mathrm{R}_1) \tag{5}$$
$$\tau(\neg\,\mathrm{R}_1, \mathrm{R}_2) = -\tau(\mathrm{R}_1, \mathrm{R}_2) \tag{6}$$
$$\tau(\text{Я}_1, \text{Я}_2) = \tau(\mathrm{R}_1, \mathrm{R}_2) \tag{7}$$
$$\tau(\neg\,\text{Я}_1, \neg\,\text{Я}_2) = \tau(\mathrm{R}_1, \mathrm{R}_2) \tag{8}$$

Proof. Equations (5) to (8) follow immediately from the definition of the τ correlation measure (see Formula 3):

(5) Symmetry of the τ measure follows from the commutativity of the max and min operators used for computing the terms of numerator as well as denominator.

(6) Negating one of the r-valued relations changes solely the sign of all r-valued equivalences in the numerator.

(7) Taking the converse relations of both r-valued relations means correspondingly transposing all (x, y) terms to (y, x) terms, jointly in numerator and denominator; thus, leaving invariant the resulting fraction.

(8) Taking the codual relations, i.e. the negation of the converse, of both r-valued relations, hence, leaves invariant their τ correlation measure.

□

In order to avoid, the case given, being fooled by randomness, we address in the next section the problem of estimating via Monte Carlo simulations the actual significance of the ordinal correlation measure when working with different types of random r-valued relations.

2 Testing for ordinal correlations

Originally, Kendall only considered correlations between crisp rankings without ties. Kendall's τ measure for pairs of random instances of such rankings of order n is known to show an expected correlation $\mu_\tau = 0.0$ with standard deviation [12]:

$$\sigma_\tau = \sqrt{\frac{2(2n+5)}{9n(n-1)}}.$$

This gives for rankings of order $n = 20$, for instance, a standard deviation $\sigma_\tau \approx 0.17$. Assuming a nearly Gaussian distribution of μ_τ, we obtain 90% and 99% confidence intervals of approximatively ± 0.22, respectively ± 0.40. Hence, a measure $|\tau| > 0.4$ observed between two rankings of 20 objects reveals a significant positive or negative ordinal correlation between them.

To similarly estimate the significance of the τ correlation measure when comparing r-valued relations, we were running extensive Monte Carlo simulations with, in turn, three specific models of random relations, namely random uniformly r-valued ones, r-valued weak tournaments and, randomly generated bipolar outranking relations resulting from the aggregation of multiple cost and benefit criteria.[10].

2.1 Randomly r-valued relations

First we consider a model of random relations where to each non reflexive pair of elements (x, y) in X is associated a uniform random float between -1.0 and 1.0. Each possible r-valued relation has thus the same probability to appear. To get a hindsight on the correlation and determinateness measures we may obtain with this genuine kind of r-valued relations, we generate large samples of 100 000 pairs of such r-valued relations for different orders n. Each pair (x, y) has, thus, in the limit, an average probability of $1/2$ to be related or not; the strict indeterminate value 0.0 having no chance to effectively appear as random number.

In Figure 1 is represented the scatter plot of the resulting tuples (d, τ) for r-valued relations of order 20. What strikes immediately is the nearly perfect symmetry of the resulting distributions, both of the determinateness degrees, as well as of the correlation measures.

In this model of random r–valued relations, the distribution of the equivalence measures E of each pairwise relational situation is following a symmetric triangular density with spread ± 1.0 and

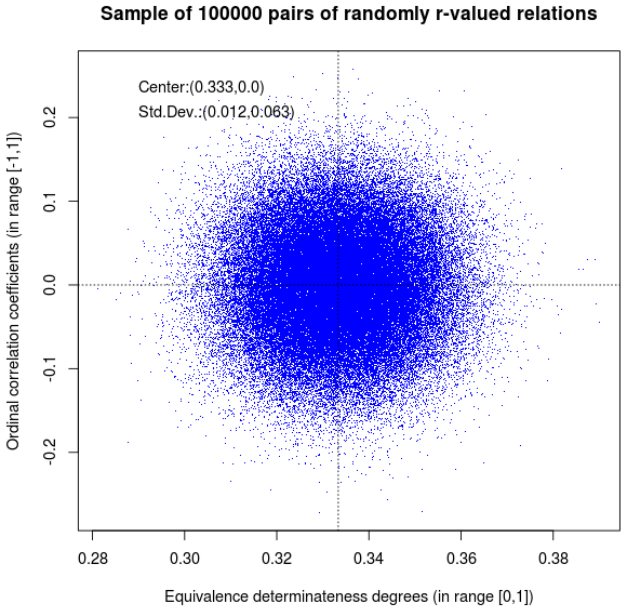

Figure 1. Scatter plot of (d, τ) for pairs of randomly r-valued relations of order 20

0 mode. Such random variables admit a mean $\mu_e = 0$ and a standard deviation $\sigma_e = \sqrt{3/18}$. A similar situation is observed when considering their equivalence determinateness measures. Each term in the denominator of Formula 3 is chosen from a same independent and identically distributed random variable D with positive density $1 - x$ for x in $[0; 1[$. This distribution – a special case of the triangular distribution where the mode equals the lower limit – shows a mean $\mu_d = 1/3$ and a standard deviation of $\sigma_d = \sqrt{1/18}$.

Hence, the observed random ordinal correlation measures $\tau = \frac{\sum E}{\sum D}$ result from the ratio of two non-independent sums of $n(n-1)$ independent and identically distributed random variables. Following from the central limit theorem, the observed statistics (see Table 3) rapidly show, with increasing order n, a more and more Gaussian distribution with mean $\hat{\mu}_\tau \approx \frac{\mu_e}{\mu_d} = 0$ and standard deviation:

$$\frac{\hat{\sigma}_\tau}{\sqrt{n(n-1)}} \approx \frac{\sigma_e}{\mu_d} \frac{1}{\sqrt{n(n-1)}}$$

getting ever smaller with increasing order n of the r-valued relations.

In Table 3 we may notice that the observed empirical standard deviations $\hat{\sigma}_d$ when multiplied with $\sqrt{n(n-1)}$ converge indeed to σ_d which equals $\sqrt{1/18} = 0.2357023$. Similarly, we may notice that the observed standard deviation $\hat{\sigma}_\tau$ tends also to the theoretical standard deviation $\sigma_\tau = 3\sqrt{3/18} = 1.224745$ when multiplied by $\sqrt{n(n-1)}$. Notice, however, a consistent negative bias of roughly 0.2%.

Example 2.1. Consider two given r-valued relations R_1 and R_2 of order 20. To test if they could have been randomly generated, we may apply a two-sided test with null hypothesis: H_0 "relations R_1 and R_2 are randomly r-valued". From the empirical results, we see that H_0 may be rejected with an error probability of 10% when $|\tau(\mathrm{R}_1, \mathrm{R}_2)| > 0.1035$ or $|d(\mathrm{R}_1, \mathrm{R}_2) - 0.3333| > 0.202$. Using, our theoretical standard deviations $\sigma_\tau = 3\sqrt{3/18}$, respectively $\sigma_d = \sqrt{1/18}$, we may precisely confirm these confidence intervals: ± 0.1033, respectively ± 0.0199 with a Gaussian test.

We have thus established a generic test apparatus for two-sided, or both positive or negative one-sided tests for measuring the significance of the ordinal correlation and equivalence determinateness of any two given r-valued relations.

Yet we are more interested in testing the correlation and equivalence determinateness when considering a specific subset of r-valued relations, namely weakly complete ones. Uniformly r-valued random relations, indeed, are statistically quite regularly structured, with on average 1/4 of double links, 1/4 of single forward links, 1/4 of single backward links and, 1/4 of no links. When working in the fields of social choice or multiple criteria decision aid with "*at least as good as*" preferential situations, we usually consider complementary concordance versus discordance relations [11] that do not allow a "*no link*" situation.

2.2 Random weakly complete relations

Formally we say that a r-valued relations R is weakly complete if for all $(x, y) \in X$, $r(x\,\mathrm{R}\,y) < 0$

Table 3. Summary Statistics, for 100000 pairs of randomly r-valued relations

$d(\mathrm{R}_1, \mathrm{R}_2)$	$\overline{d}$	$\hat{\sigma}_d$	$\hat{\sigma}_d\sqrt{n(n-1)}$	Conf. 90%	Conf. 99%
$n = 5$	0.3333	0.0527	0.23568	±0.0866	±0.1355
$n = 10$	0.3334	0.0249	0.23622	±0.0406	±0.0645
$n = 15$	0.3333	0.0162	0.23476	±0.0266	±0.0418
$n = 20$	0.3333	0.0121	0.23587	±0.0202	±0.0276
$n = 30$	0.3333	0.0080	0.23597	±0.0132	±0.0207
$n = 50$	0.3333	0.0048	0.23758	±0.0078	±0.0121
$\tau(\mathrm{R}_1, \mathrm{R}_2)$	$\overline{\tau}$	$\hat{\sigma}_\tau$	$\hat{\sigma}_\tau\sqrt{n(n-1)}$	Conf. 90%	Conf. 99%
$n = 5$	0.0003	0.2731	1.22134	±0.4500	±0.6766
$n = 10$	0.0000	0.1289	1.22285	±0.2181	±0.3291
$n = 15$	0.0000	0.0842	1.22017	±0.1386	±0.2156
$n = 20$	0.0000	0.0621	1.21055	±0.1035	±0.1425
$n = 30$	0.0000	0.0414	1.22113	±0.0681	±0.1064
$n = 50$	0.0000	0.0247	1.22259	±0.0406	±0.0636

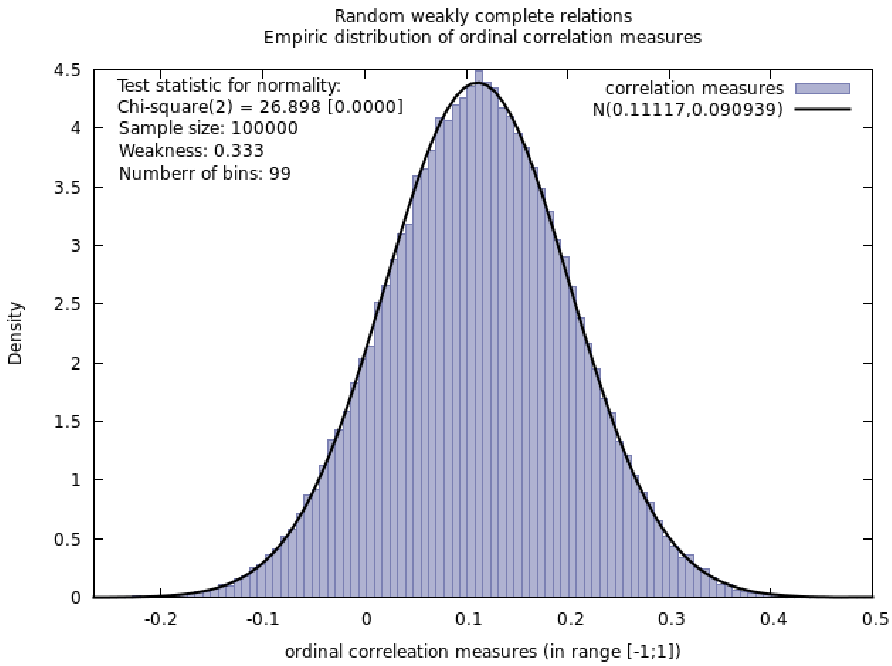

Figure 2. Histogram of correlation measures with normality test for pairs of random weakly complete relations of order 20

implies $r(x\,\mathrm{R}\,y) \geq 0$. Each link is therefore either a double, or a single forward or backward link, each one with equal probability 1/3.

Determinateness distribution of equivalence of pairs of this model of random r-valued relations remains very close to 1/3, as in the general model above (see Table 4), except a slight lowering of its mean values (compare with Table 3). Similarly, we may again observe an empiric distribution of correlation measures which follows, with increasing order of the relations, more and more, due to the central limit theorem, a Gaussian distribution. In Figure 2 is represented a histogram from a sample of 100 000 random instances of weakly complete r-valued relations of order 20. Notice first the fact that the sampled mean correlation measure $\hat{\mu}_\tau$ is shifted roughly by +0.111, depending on the given degree of weakness. In the limit, a weakness degree of 1.0, on the one hand, would give always the same complete relation, showing, hence, a constant correlation measure of 1.0. A weakness degree of 0.0, on the other hand, would give samples of random tournaments with mean and median correlation measures concentrated around 0 as in the general case above.

In Table 4 we summarize empiric statistical results for weakly complete r-valued relations of different orders, maintaining constant a weakness degree of 1/3. The observed distribution of correlation measures τ, besides the already mentioned positive shift of the mean by approximately +0.111, also shows an empiric standard deviation $\hat{\sigma}_\tau$ multiplied by $\sqrt{n(n-1)}$ that is no longer a constant independent of the given order n. Hence,

Table 4. Summary Statistics, for 100000 pairs of random weakly (1/3) complete relations

$d(\mathrm{R}_1, \mathrm{R}_2)$	$\hat{\mu}_d$	$\hat{\sigma}_d$	$\hat{\sigma}_d\sqrt{n(n-1)}$	Conf. 90%		Conf. 99%	
$n = 5$	0.33344	0.05268	0.23568	0.24920	0.42208	0.20493	0.47489
$n = 10$	0.33316	0.02490	0.23622	0.29262	0.37433	0.27069	0.39861
$n = 15$	0.33316	0.01634	0.23476	0.30646	0.36025	0.29164	0.37578
$n = 20$	0.33320	0.01209	0.23587	0.31342	0.35315	0.30262	0.36486
$n = 30$	0.33316	0.00799	0.23597	0.32006	0.34630	0.31269	0.35406
$n = 50$	0.33318	0.00477	0.23758	0.32537	0.34102	0.32099	0.34558
$\tau(\mathrm{R}_1, \mathrm{R}_2)$	$\hat{\mu}_\tau$	$\hat{\sigma}_\tau$	$\hat{\sigma}_\tau\sqrt{n(n-1)}$	Conf. 90%		Conf. 99%	
$n = 5$	0.1112	0.3032	1.3560	−0.3981	+0.6039	−0.6559	+0.8229
$n = 10$	0.1113	0.1592	1.5103	−0.1537	+0.3713	−0.3019	+0.5082
$n = 15$	0.1112	0.1138	1.6491	−0.0767	+0.2978	−0.1810	+0.3978
$n = 20$	0.1112	0.0909	1.7720	−0.0395	+0.2604	−0.1231	+0.3399
$n = 30$	0.1116	0.0681	2.0087	−0.0003	+0.2234	−0.0631	+0.2869
$n = 50$	0.1112	0.0484	2.3957	+0.0313	+0.1905	−0.0132	+0.2348

the equivalence measures on the numerator of the τ measures are no longer independent and identically distributed. Consequently, the central limit theorem is no longer automatically applicable. When verifying the plausibility of the randomness hypothesis when comparing weakly complete r-valued relations, we are thus solely left with the potentially biased sample standard deviations and the corresponding tail percentiles estimations.

Example 2.2. With 99 bins, the χ^2 test (see Figure 2), however, clearly confirms ($26.898 \ll \chi^2(0.01, 98) = 68.396$, p-value $= 0.0$) for order $n = 20$ and average weakness 1/3, the quality of the Gaussian approximation with empirical mean $\hat{\mu}_\tau = 0.1112$ and standard deviation $\hat{\sigma}_\tau = 0.0909$. The Gaussian 90%-confidence, resp. 99%-confidence, interval of the mean correlation measure μ_τ, hence, gives the limits $[-0.0384; +0.2606]$, respectively $[-0.1231; +0.3454]$. And, both intervals are, indeed, very close to the empirical ones (see Table 4, row $n = 20$) we obtain with a sample of 100 000 random instances.

Finally, we consider a special subset of weakly complete r-valued relations, namely r-valued *outranking* relations.

2.3 Random outranking relations

Concordance relations, i.e. weakly complete r-valued relations naturally result from the ordinal aggregation of multiple performance criteria when considering the weighted concordance of the statements: "*x performs at least as good as y*" [10]. Our random model for such kind of r-valued relations is based on randomly generated performances for all decision actions in $x \in X$ on each criterion. We distinguish three types of decision actions: *cheap*, *neutral* and *expensive* ones with an equal proportion of 1/3. We also distinguish two types of weighted criteria: *cost* criteria to be *minimized*, and *benefit* criteria to be *maximized*; in the proportions 1/3 respectively 2/3. Random performances on each type of criteria are drawn, either from an ordinal scale [0; 10], or from a cardinal scale [0.0; 100.0], following a parametric triangular law of mode: 30% performance for cheap, 50% for neutral, and 70% performance for expensive decision actions, with constant probability repartition 0.5 on each side of the respective mode. Cost criteria use mostly cardinal scales (3/4), whereas benefit criteria use mostly ordinal scales (2/3). The sum of weights of the cost criteria always equals the sum weights of the benefit criteria. On cardinal criteria, both of cost or of benefit type, we observe following constant preference discrimination quantiles: 5% indifferent situations, 90% strict preference situations 90%, and 5% veto situation. We call this random model of r-valued relations for short random *CB-outranking* relations.

In Table 5 we summarize the empirical results for various numbers of decision actions (n) and criteria (c). Most noticeable is here the diminishing average determination degrees with rising numbers n of actions and, especially numbers c of criteria. Indeed, the fixed proportion of veto situations (5%) on each cardinal criteria augments, with the number of criteria, the probability of the presence of pairwise indeterminate, i.e. 0-valued, outranking situations. Furthermore, the empiric distribution of the determination degrees appears no more to converge to a Gaussian type limit.

In Figure 3, one may notice, indeed, on a sample of 100 000 random CB-outranking relations of order $n = 20$ and criteria $c = 13$, an apparent left asymmetry, confirmed by a positive skewness of 0.876, as well as a clearly leptocurtic distribution (excess kurtosis: +1.7884) of the observed determinateness degrees. Comparing the observed distribution with a theoretical gamma distribution, reveals a positive match with parameters: $\alpha = 38.119$ and $\beta = 0.004$. With order $n = 30$ and criteria $c = 21$ one obtains a similar gamma estimation with parameters: $\alpha = 64.594$ and $\beta = 0.002$.

With rising numbers of indeterminate preferential situations, the proportion of double links compared to single links, is no more as regular (1/3 against 2/3) as in the genuine model of random weakly complete relations. In the $n = 20$ and

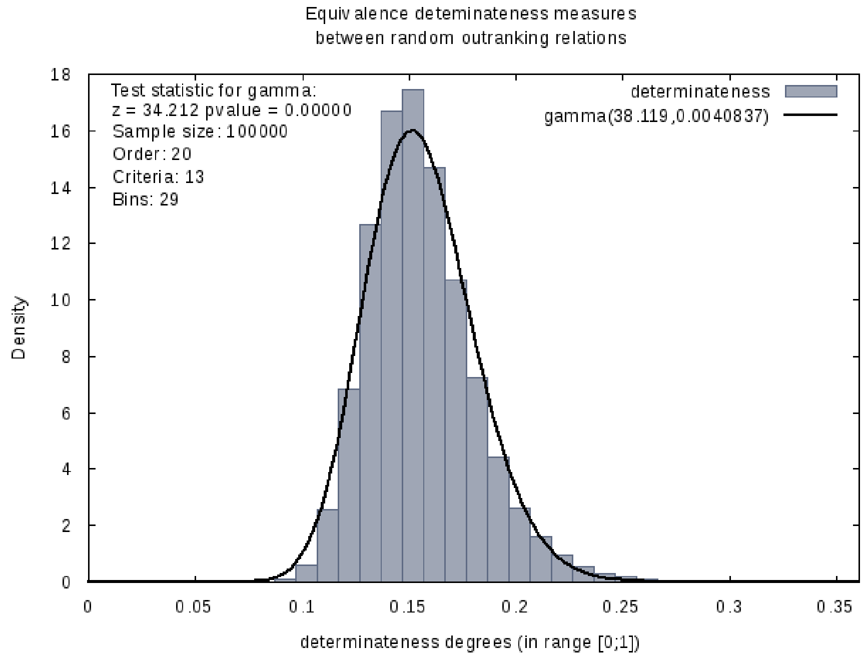

Figure 3. Histogram of determination degrees for pairs of random CB-outranking relations of order 20

Table 5. Summary Statistics, for 100000 pairs of random CB-outranking relations

$d(\mathrm{R}_1, \mathrm{R}_2)$	$\hat{\mu}_d$	$\hat{d}_{50\%}$	$\hat{\sigma}_d$	Conf. 90%		Conf. 99%	
$n = 5, c = 3$	0.3259	0.3250	0.1131	0.1500	0.5255	0.0750	0.6333
$n = 10, c = 7$	0.2207	0.2165	0.0482	0.1494	0.3072	0.1204	0.3681
$n = 15, c = 9$	0.1910	0.1867	0.0362	0.1399	0.2577	0.1196	0.3102
$n = 20, c = 13$	0.1557	0.1527	0.0252	0.1203	0.2013	0.1053	0.2435
$n = 30, c = 21$	0.1372	0.1357	0.0174	0.1120	0.1674	0.1002	0.1989
$\tau(\mathrm{R}_1, \mathrm{R}_2)$	$\hat{\mu}_\tau$	$\hat{\tau}_{50\%}$	$\hat{\sigma}_\tau$	Conf. 90%		Conf. 99%	
$n = 5, c = 3$	0.0378	0.0345	0.5145	−0.7929	+0.8610	−1.0000	+1.0000
$n = 10, c = 7$	0.0629	0.0644	0.3037	−0.4420	+0.5560	−0.6483	+0.7467
$n = 15, c = 9$	0.0727	0.0761	0.2417	−0.3323	+0.4667	−0.5206	+0.6354
$n = 20, c = 13$	0.0984	0.1017	0.2085	−0.2492	+0.4383	−0.4224	+0.5904
$n = 30, c = 21$	0.1239	0.1272	0.1712	−0.1639	+0.4007	−0.3162	+0.5339

$c = 13$ case, with a sample of 100 000 random instances, we observe only 18.7% double links, with 71.6% single links and, in this particular case, 9.8% of indeterminate links. Consequently, in Table 5, we observe lower average correlation measures $\hat{\mu}_\tau$ than with the previous model. Furthermore, the convergence to a Gaussian limit distribution with rising order of the relations is no more apparent when considering in Figure 4 the Q-Q plot of the simulated correlation quantiles against Gaussian quantiles for the case $n = 20$ and $c = 13$. A very high χ^2 value (567.356) rejects, indeed, the Gaussian approximation hypothesis.

Appreciating the significance of the correlation between pairs of CB-outranking relations remains, hence, solely possible on the basis of sampled tail percentiles from Monte Carlo simulations. In the appendix we have gathered estimated 5%, 95%, 0.5% and 99.5% percentiles for relations of various orders and numbers of criteria that may be relevant in an MCDA context.

Example 2.3. Let us eventullly consider the random r-valued CB-outranking relation shown in Table 6. Relation R_1, with an average determination degree $d(R_1) = 0.397$, is defined on $n = 10$ decision actions and results from the ordinal concordance observed on $c = 7$ performance criteria. Applying for instance Kemeny's ranking rule [13] would give us the following crisp linear order: $[4, 2, 7, 8, 9, 1, 10, 6, 3, 5]$, showing a highly significant correlation of $+0.888$ with R_1 (see the upper limit 0.747 of the 99% confidence interval in Table 5). Indeed, under the hypothesis of a completely random ordering, such a high correlation measure would appear in less than 0.5% of cases. When ranking now with the help of the net flows scores à la PROMETHEE [14], we would obtain the order: $[4, 9, 2, 7, 8, 10, 1, 6, 3, 5]$, showing a less higher correlation $(+0.776)$ with R_1.

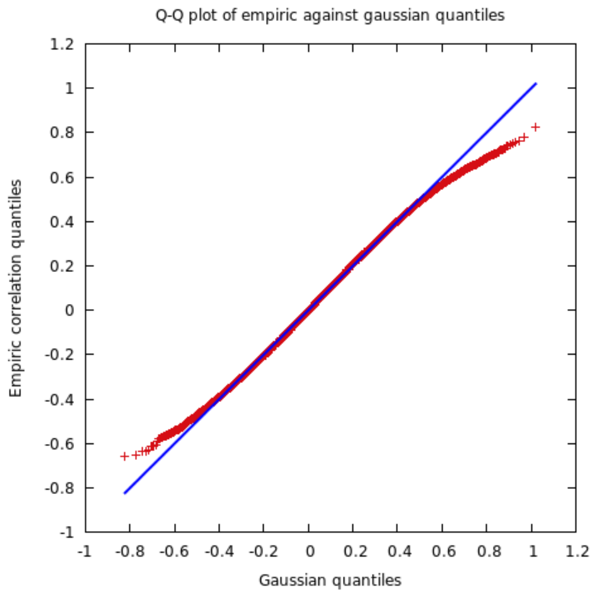

Figure 4. Q-Q plot of empiric again normal correlation quantiles for pairs of random CB-outranking relations of order 20

Table 6. Example of random CB-outranking relation ($n = 10$, $c = 7$)

R_1	1	2	3	4	5	6	7	8	9	10
1	–	$+0.14$	$+0.43$	-0.14	$+0.29$	$+0.14$	$+0.43$	-0.14	± 0.00	$+0.43$
2	$+0.43$	–	$+0.43$	-0.43	$+0.43$	$+0.14$	$+0.14$	$+0.43$	$+0.14$	$+0.14$
3	-0.43	-0.43	–	-0.71	$+0.43$	$+0.00$	-0.43	-1.00	-1.00	-0.14
4	$\mathbf{+0.14}$	$\mathbf{+0.71}$	$\mathbf{+1.00}$	–	$\mathbf{+0.71}$	$\mathbf{+0.43}$	$\mathbf{+0.71}$	$\mathbf{+0.43}$	$\mathbf{+0.14}$	$\mathbf{+0.57}$
5	$+0.14$	-0.43	-0.43	-0.71	–	-0.71	-1.00	$+0.14$	-1.00	-0.43
6	-0.14	-0.14	$+1.00$	-0.43	$+0.71$	–	-0.14	-0.14	$+0.14$	-0.43
7	$+0.14$	$+0.14$	$+0.43$	-0.43	$+1.00$	$+0.14$	–	$+0.14$	$+0.43$	$+0.29$
8	$+0.43$	-0.14	$+1.00$	-0.43	$+0.43$	$+0.14$	-0.14	–	$+0.43$	-0.14
9	$+1.00$	-0.14	$+1.00$	-0.14	$+1.00$	$+0.43$	$+0.14$	-0.14	–	$+0.14$
10	-0.43	-0.14	$+0.43$	$+0.14$	$+0.43$	$+0.43$	$+0.29$	$+0.14$	-0.14	–

Kohler's rule would, furthermore, give us the order: $[4, 2, 8, 10, 9, 6, 1, 7, 3, 5]$ with the same correlation of $+0.776$. Finally, Tideman's ranked pairs rule [15], in fact the dual of Dias and Lamboray's leximin rule [3], will deliver the order: $[4, 2, 8, 9, 1, 7, 10, 6, 3, 5]$ with, this time again, a highly significant correlation measure of $+0.872$. As we compare here each time R_1 with a 1.0-valued (crisp) linear order, the correlation underlying equivalence determination degree is actually, in all cases, equal to $d(R_1) = 0.397$. By the way, we may notice that the same first ranked decision action with all the ranking rules is action 4, in fact a CONDORCET winner that outranks all other decision actions with a majority of at least 57% of the criteria weights (see row 4 in Table 6).

Conclusion

We have consistently generalized Kendall's rank correlation measure τ to r-valued binary relations via a corresponding r-valued logical equivalence measure. The so extended ordinal correlation measure, besides remaining identical to Kendall's measure in the case of completely determined linear orders, shows interesting properties like its independence with the actual determinateness degree of the r-valued equivalence. Empirical confidence intervals for different models of random r-valued relations, like weakly complete and, more particularly, r-valued outranking relations are elaborated.

References

[1] Bisdorff R (2008) On clustering the criteria in an outranking based decision aid approach. In Le Thi H A et al. (eds) Modelling, Computation and Optimization in Information Systems and Management Sciences Springer-Verlag CCIS 14:409–418 1

[2] Bisdorff R and Meyer P and Olteanu A (2011) A Clustering Approach using Weighted Similarity Majority Margins. In Tang J et al (Eds) Advanced Data Mining and Applications ADMA 2011 Part I Springer-Verlag LNAI 7120:15-28 1

[3] Dias L C and Lamboray C (2010) Extensions of the prudence principle to exploit a valued outranking relation. European Journal of Operational Research 201(3):828–837 1, 9

[4] Ruiz D and Hüllermeier E (2012) A formal and empirical analysis of the gamma rank correlation coefficient. Information Sciences 206:1–17 1

[5] Bisdorff R (2000) Logical foundation of fuzzy preferential systems with application to the electre decision aid methods. Computers and Operations Research 27:673-687 1, 2

[6] Bisdorff R (2002) Logical Foundation of Multicriteria Preference Aggregation. In: Bouyssou D et al (eds) Essay in Aiding Decisions with Multiple Criteria. Kluwer Academic Publishers 379–403 1, 2

[7] Kendall M G (1938) A New Measure of Rank Correlation. Biometrica 30:81–93 1

[8] Kendall M G (1955) Rank Correlation Methods. Hafner Publishing Co New York 1

[9] Degenne A (1972) Techniques ordinales en analyse des données statistique. Hachette Paris 1

[10] Bisdorff R (2012) On polarizing outranking relations with large performance differences. Journal of Multi-Criteria Decision Analysis DOI: 10.1002/mcda.1472 1–20 2, 4, 7

[11] Bouyssou D and Pirlot M (2005) A characterization of concordance relations. European Journal of Operational Research, 167(2):427–443 5

[12] Valz P D and McLeod A I (1990) A Simplified Derivation of the Variance of Kendall's Rank Correlation Coefficient. The American Statistician 44(1) 39–40 4

[13] Lamboray C (2007) A comparison between the prudent order and the ranking obtained with Borda's, Copeland's, Slater's and Kemeny's rules. Mathematical Social Sciences 54:1–16 8

[14] Brans JP and Vincke Ph (1985) A preference ranking organisation method : The PROMETHEE method for MCDM. Management Science 31(6):647–656 1, 8

[15] Lamboray C (2009) A prudent characterization of the Ranked Pairs Rule. Social Choice and Welfare 32:129–155 9