



Universiteit
Leiden
The Netherlands

Kappa coefficients for circular classifications

Warrens, M.J.; Bunga Citra Pratiwi, B.

Citation

Warrens, M. J., & Bunga Citra Pratiwi, B. (2016). Kappa coefficients for circular classifications. *Journal Of Classification*, 33(3), 507-522. doi:10.1007/s00357-016-9217-3

Version: Not Applicable (or Unknown)

License: [Leiden University Non-exclusive license](#)

Downloaded from: <https://hdl.handle.net/1887/45185>

Note: To cite this publication please use the final published version (if applicable).

Kappa Coefficients for Circular Classifications

Matthijs J. Warrens & Bunga C. Pratiwi

Journal of Classification

ISSN 0176-4268

Volume 33

Number 3

J Classif (2016) 33:507-522

DOI 10.1007/s00357-016-9217-3



Your article is published under the Creative Commons Attribution license which allows users to read, copy, distribute and make derivative works, as long as the author of the original work is cited. You may self-archive this article on your own website, an institutional repository or funder's repository and make it publicly available immediately.

Kappa Coefficients for Circular Classifications

Matthijs J. Warrens

University of Groningen, The Netherlands

Bunga C. Pratiwi

Leiden University, The Netherlands

Abstract: Circular classifications are classification scales with categories that exhibit a certain periodicity. Since linear scales have endpoints, the standard weighted kappas used for linear scales are not appropriate for analyzing agreement between two circular classifications. A family of kappa coefficients for circular classifications is defined. The kappas differ only in one parameter. It is studied how the circular kappas are related and if the values of the circular kappas depend on the number of categories. It turns out that the values of the circular kappas can be strictly ordered in precisely two ways. The orderings suggest that the circular kappas are measuring the same thing, but to a different extent. If one accepts the use of magnitude guidelines, it is recommended to use stricter criteria for circular kappas that tend to produce higher values.

Keywords: Cohen's kappa; Weighted kappa; Inter-rater agreement; Linear scale; Circular scale.

The authors thank two anonymous referees for their helpful comments and valuable suggestions on an earlier version of this article.

Corresponding Author's Address: Matthijs J. Warrens, GION, University of Groningen, Grote Rozenstraat 3, 9712 TG Groningen, The Netherlands, e-mail: m.j.warrens@rug.nl.

Published online: 11 November 2016

1. Introduction

Similarity coefficients are used in pattern recognition, data analysis and classification to quantify the strength of a relationship between two variables or classifications. Similarity coefficients can be used to summarize parts of a research study, but can also be used as input for data-analytic techniques, for example, cluster analysis. Well-known examples of similarity coefficients are Pearson's product-moment correlation for measuring linear dependence between two interval variables, the Jaccard coefficient for measuring co-occurrence of two species types, and the Hubert-Arabie adjusted Rand index for comparing partitions obtained with different clustering algorithms (Warrens 2008, 2014). Kappa coefficients are commonly used to quantify agreement between classifications with identical categories (Vanbelle, Mutsvari, Declerck, and Lesaffre 2012; Warrens 2010a, 2011a; Yang and Zhou 2015).

In social and behavioral science and biomedical research, it is frequently required that agreement between two classifications with identical categories is assessed. For example, to assess the reliability of a rating scale researchers typically let two observers rate independently the same set of objects. The categories of the rating scale are defined in advance. The agreement between the observers can be used to investigate the reliability of the rating scale. Standard tools for quantifying agreement between classifications with identical categories are Cohen's kappa in the case of nominal categories (Yang and Zhou 2014; Warrens 2010b), and weighted kappa in the case of ordinal categories (Vanbelle 2015; Yang and Zhou 2015; Warrens 2012, 2013, 2015). Both coefficients correct for agreement due to chance.

Although interval and ordinal data are usually measured on a linear scale, data may also exhibit a certain periodicity, for example, if the data are naturally measured on a circular scale. Examples of circular interval data are directions measured in degrees, and the time of the day (Berens 2009). Examples of categorical classifications that have been measured on a circular scale are the day of the week, affect states (Posner, Russell, and Peterson 2005; Watson, Wiese, Vaidya, and Tellegen 1999; Watson and Tellegen 1985), vocational interests (Brown 1992), and phases of cell cycle genes (Rueda, Fernández, and Peddada 2009). In social and behavioral science and biomedical research, circular scales that measure social or psychological constructs are usually referred to as circumplex models.

With circular scales the designation of high and low is arbitrary. Furthermore, with categorical circular scales an anchor point is usually not appropriate. For example, Russell (1980) hypothesized that the following eight affect categories can be ordered on a circular scale: Arousal, Excitement, Pleasure, Contentment, Sleepiness, Depression, Misery and Distress.

Table 1. Hypothetical pairwise classifications of 200 photos with facial expressions into eight affect categories by two human classifiers.

Classifier 1	Classifier 2								Total
	A ₁	A ₂	A ₃	A ₄	A ₅	A ₆	A ₇	A ₈	
A ₁ = Arousal	24	3	0	0	0	0	0	2	29
A ₂ = Excitement	2	16	1	0	0	0	0	0	19
A ₃ = Pleasure	0	1	15	3	0	0	0	0	19
A ₄ = Contentment	0	0	4	13	5	0	0	0	22
A ₅ = Sleepiness	0	0	0	2	18	3	0	0	23
A ₆ = Depression	0	0	0	0	4	22	3	0	29
A ₇ = Misery	0	0	0	0	0	3	26	3	32
A ₈ = Distress	3	0	0	0	0	0	2	22	27
Total	29	20	20	18	27	28	31	27	200

Table 1 presents hypothetical pairwise classifications of 200 photos with facial expressions into Russell’s eight affect categories by two classifiers. Because the categories of the rows and columns of Table 1 are in the same order, the elements on the main diagonal are the number of photos on which the classifiers agreed. All off-diagonal elements are numbers of photos on which the classifiers disagreed. With the depicted ordering of the rows and columns of Table 1, there is only disagreement between the classifiers on adjacent categories. The disagreement between Arousal and Distress suggests that the two affect states should be adjacent on a scale. The categories in Table 1 thus form a circular scale. More elaborate circular scales of affect states can be found in Posner, Russell, and Peterson (2005) and Watson, Wiese, Vaidya, and Tellegen (1999).

A second example comes from career assessment. Assessment of vocational interest is done to give insight into a person’s interests, so that participants may be assisted in the choice of an occupation that will sustain their interests and keep them usefully employed throughout their working life. Vocational interest is usually measured with an interest inventory. A participant who completes an interest inventory expresses preferences about items concerning a field of work or recreation. The outcome of an interest inventory is one or a combination of the following six ordered categories: Realistic, Investigative, Artistic, Social, Enterprising and Conventional.

Table 2 presents hypothetical pairwise classifications of the primary vocational interest of 120 participants obtained with two different interest inventories. The elements on the main diagonal are the numbers of participants with the same vocational interest according to both inventories. All off-diagonal elements are numbers of participants on which the inventories disagreed. Most disagreement is on categories that are adjacent in the depicted ordering. Furthermore, the disagreement between Realistic and

Table 2. Hypothetical pairwise classifications of the primary vocational interest of 120 participants into six categories by two different interest inventories.

Inventory 1	Inventory 2						Total
	A_1	A_2	A_3	A_4	A_5	A_6	
$A_1 = \text{Realistic}$	12	2	1	0	1	2	18
$A_2 = \text{Investigative}$	2	13	1	2	0	1	19
$A_3 = \text{Artistic}$	1	1	8	3	0	0	13
$A_4 = \text{Social}$	0	1	2	17	5	0	25
$A_5 = \text{Enterprising}$	1	0	1	2	9	3	16
$A_6 = \text{Conventional}$	2	2	0	1	2	22	29
Total	18	19	13	25	17	28	120

Conventional and their adjacent categories suggests that the two categories should be adjacent on a scale. The categories in Table 2 thus form a circular scale.

The standard weighted kappas for linear scales studied in, for example, Vanbelle (2015) and Warrens (2012, 2013, 2014, 2015), are not appropriate with circular scales since they require that the scale has endpoints, which is not the case with circular scales. A kappa coefficient for circular classifications with identical categories as a special case of weighted kappa was first presented in Gwet (2012, p. 63-64). In an example with four categories, Gwet suggested to assign weights only to agreement, and disagreements on adjacent categories. In this paper, we generalize this idea and define formally a family of kappa coefficients for circular classifications. Furthermore, it is shown how the circular kappas are related, and it is studied whether the circular kappas depend on the number of categories.

The paper is organized as follows. In Section 2, we introduce the notation and present several definitions. A family of circular kappas is defined in Section 3. In Section 4, it is shown that the circular kappas can be ordered in two ways. One ordering is more likely to occur in practice. The second ordering is the reverse ordering of the first one. Furthermore, it is shown that a specific class of circular kappas can be interpreted as weighted averages of the Cohen’s kappas of all collapsed tables that are obtained by combining two adjacent categories. In Section 5, a possible dependence of the circular kappas on the number of categories is studied. A result is presented that suggests that the circular kappas tend to increase with the number of categories. A discussion and several recommendations are presented in Section 6.

2. Notation and Weighted Kappa

Suppose that two fixed classifiers (for example, expert observers, algorithms, rating instruments) have independently classified the same set

of n objects (for example, individuals, scans, products) into the categories A_1, A_2, \dots, A_c , that were defined in advance. For a population of objects, let π_{ij} denote the proportion of the n objects that is classified into category A_i by the first classifier and into category A_j by the second classifier, where $i, j \in \{1, 2, \dots, c\}$. We assume that the categories of the rows and columns of the table $\{\pi_{ij}\}$ are in the same order, so that the diagonal elements π_{ii} reflect the exact agreement between the two classifiers. In the context of agreement studies, the table $\{\pi_{ij}\}$ is usually called an agreement table.

Define

$$\pi_{i+} := \sum_{j=1}^c \pi_{ij}, \tag{1}$$

and

$$\pi_{+i} := \sum_{j=1}^c \pi_{ji}. \tag{2}$$

The marginal probabilities π_{i+} and π_{+i} reflect how often the categories were used by the first and second classifier, respectively. Furthermore, if the ratings between the two classifiers are statistically independent the expected value of π_{ij} is given by $\pi_{i+}\pi_{+j}$. The table $\{\pi_{i+}\pi_{+j}\}$ contains the expected values.

In the next section, we define kappa coefficients for circular classifications as special cases of weighted kappa. Weighted kappa is a standard tool for quantifying the degree of agreement between two classifications with ordinal categories. With ordered categories, there is usually more disagreement between the classifiers on adjacent categories than on categories that are further apart. Weighted kappa allows the user to describe the closeness between categories using weights (Vanbelle 2015; Warrens 2013, 2014). The real number $0 \leq w_{ij} \leq 1$ denotes the weight corresponding to cell (i, j) of tables $\{\pi_{ij}\}$ and $\{\pi_{i+}\pi_{+j}\}$. The weighted kappa coefficient is defined as (Warrens 2011b)

$$\kappa = \frac{\sum_{i=1}^c \sum_{j=1}^c w_{ij} (\pi_{ij} - \pi_{i+}\pi_{+j})}{1 - \sum_{i=1}^c \sum_{j=1}^c w_{ij} \pi_{i+}\pi_{+j}}. \tag{3}$$

The cell probabilities of the table $\{\pi_{ij}\}$ are not directly observed. Let $\{n_{ij}\}$ denote the contingency table of observed frequencies. Assuming a multinomial sampling model with the total number of objects n fixed, the maximum likelihood estimate of π_{ij} is given by $\hat{\pi}_{ij} = n_{ij}/n$ (Yang and Zhou 2014, 2015). Tables 1 and 2 are examples of $\{n_{ij}\}$. Furthermore, under the multinomial sampling model, the maximum likelihood estimate of κ is

$$\hat{\kappa} = \frac{\sum_{i=1}^c \sum_{j=1}^c w_{ij} (n_{ij}/n - n_{i+}n_{+j}/n^2)}{1 - \sum_{i=1}^c \sum_{j=1}^c w_{ij} n_{i+}n_{+j}/n^2}. \tag{4}$$

Estimate (4) is obtained by substituting $\hat{\pi}_{ij} = n_{ij}/n$ for the cell probabilities π_{ij} in (3). Furthermore, a large sample standard error for weighted kappa is presented in Fleiss, Cohen, and Everitt (1969) (see also, Yang and Zhou 2015). This formula will be used to estimate 95% confidence intervals of the point estimate $\hat{\kappa}$ (see Table 3).

Next, we define several quantities for notational convenience. Consider the table $\{\pi_{ij}\}$ with relative frequencies, and define the quantities

$$\lambda_0 := \sum_{i=1}^c \pi_{ii}, \tag{5a}$$

$$\lambda_1 := \sum_{i=1}^{c-1} (\pi_{i,i+1} + \pi_{i+1,i}) + \pi_{1c} + \pi_{c1}, \tag{5b}$$

$$\lambda_2 := 1 - \lambda_0 - \lambda_1. \tag{5c}$$

Quantity λ_0 is the total observed agreement, the proportion of objects that have been classified into the same categories by both classifiers. Quantity λ_1 is the sum of the elements on the first diagonal above the main diagonal of the table $\{\pi_{ij}\}$ and the first diagonal below the main diagonal, together with the elements π_{1c} and π_{c1} . Quantity λ_1 is the proportion of disagreement on adjacent categories of the circular scale. Since $1 - \lambda_0$ is the total disagreement, quantity λ_2 is composed of the disagreement that is not part of λ_1 .

Next, consider the table $\{\pi_{i+}\pi_{+j}\}$, and define the quantities

$$\mu_0 := \sum_{i=1}^c \pi_{i+}\pi_{+i}, \tag{6a}$$

$$\mu_1 := \sum_{i=1}^{c-1} (\pi_{i+}\pi_{+(i+1)} + \pi_{(i+1)+}\pi_{+i}) + \pi_{1+}\pi_{+c} + \pi_{c+}\pi_{+1}, \tag{6b}$$

$$\mu_2 := 1 - \mu_0 - \mu_1. \tag{6c}$$

Quantities μ_0 , μ_1 and μ_2 are the expected values of quantities λ_0 , λ_1 and λ_2 , respectively, under statistical independence of the classifiers.

3. Circular Kappas

In this section, we define a family of kappa coefficients that can be used for quantifying agreement between two circular classifications. The

Table 3. Point and interval estimates of the circular kappas in (10) ($u = 0.00, 0.25, 0.50, 0.75$) for the data in Tables 1 and 2.

	Value of u	Point estimate	95% Confidence interval
Table 1	0.00	0.75	0.68 – 0.81
	0.25	0.80	0.74 – 0.85
	0.50	0.85	0.81 – 0.89
	0.75	0.92	0.90 – 0.94
Table 2	0.00	0.61	0.50 – 0.71
	0.25	0.64	0.54 – 0.73
	0.50	0.68	0.59 – 0.77
	0.75	0.73	0.64 – 0.82

kappas differ only by one parameter. Let $0 \leq u < 1$ be a real number. The number u will be used as a parameter to assign weight to the disagreement on adjacent categories. Similar to the small example in Gwet (2012), we will give full weight to the entries on the main diagonal of $\{\pi_{ij}\}$, and a partial weight u to the entries corresponding to adjacent categories; all other weights are set to 0:

$$w_{ij} := \begin{cases} 1, & \text{if } i = j; \\ u, & \text{if } |i - j| = 1, \text{ or } |i - j| = c - 1; \\ 0, & \text{otherwise.} \end{cases} \tag{7}$$

This weighting scheme makes sense if we expect some disagreement between the classifiers on adjacent categories but no serious disagreement on categories that are further apart on the scale.

Using the quantities in (5), the weighted observed agreement with parameter u is defined as

$$O_u := \lambda_0 + u\lambda_1. \tag{8}$$

Furthermore, using the quantities in (6) the expected value of (8) under independence is given by

$$E_u := \mu_0 + u\mu_1. \tag{9}$$

By using higher values of u in (8) and (9) more weight is given to the total disagreement between adjacent categories. Using (8) and (9) a family of circular kappas with parameter u can be defined as

$$\kappa_u := \frac{O_u - E_u}{1 - E_u} = \frac{\lambda_0 + u\lambda_1 - \mu_0 - u\mu_1}{1 - \mu_0 - u\mu_1}. \tag{10}$$

The value of (10) is equal to 1 if there is perfect agreement between the classifiers ($\lambda_0 = 1$), and 0 when $\lambda_0 + u\lambda_1 = \mu_0 + u\mu_1$. Formula (10) is also obtained if one uses weighting scheme (7) in the general formula (3).

For $u = 0$ we obtain Cohen's kappa (Yang and Zhou 2014; Warrens 2010b)

$$\kappa_0 = \frac{\lambda_0 - \mu_0}{1 - \mu_0} = \frac{\sum_{i=1}^c (\pi_{ii} - \pi_{i+} \pi_{+i})}{1 - \sum_{i=1}^c \pi_{i+} \pi_{+i}}. \quad (11)$$

This is an important special case of (10). The value of (11) is equal to 1 when there is perfect agreement between the classifiers ($\lambda_0 = 1$), 0 when the observed agreement is equal to that expected under independence ($\lambda_0 = \mu_0$), and negative when agreement is less than expected by chance.

Table 3 presents point and interval estimates of (10) for the data in Tables 1 and 2, and for four values of u . For example, for Table 1 the estimate of Cohen's kappa is $\hat{\kappa}_0 = 0.75$ with 95% CI = 0.68 – 0.81. The values in Table 3 illustrate that the value of the circular kappas in (10) are increasing in the parameter u for the data in Tables 1 and 2. This property is formally proved in Lemma 2 in the next section.

We end this section with the following result. Lemma 1 shows that all special cases of (10) coincide with $c = 3$ categories (and thus also with $c = 2$ categories). More precisely, Lemma 1 shows that with $c = 3$ categories all circular kappas coincide with Cohen's kappa in (11).

Lemma 1. *If $c = 3$, then $\kappa_u = \kappa_0$.*

Proof. With $c = 3$ categories, we have $\lambda_2 = 0$ and $\mu_2 = 0$, and thus the identities $\lambda_1 = 1 - \lambda_0$ and $\mu_1 = 1 - \mu_0$. Using these identities in (10) we obtain

$$\kappa_u = \frac{\lambda_0 + u(1 - \lambda_0) - \mu_0 - u(1 - \mu_0)}{1 - \mu_0 - u(1 - \mu_0)} = \frac{(1 - u)\lambda_0 - (1 - u)\mu_0}{1 - u - (1 - u)\mu_0}. \quad (12)$$

Dividing all terms on the right-hand side of (12) by $1 - u$ yields Cohen's kappa in (11).

■

Cohen's kappa is a standard tool for quantifying agreement between two classifications with nominal categories (Yang and Zhou 2014; Warrens 2010b). Lemma 1 shows that in the case of three categories the kappa coefficients for nominal categories and circular categories coincide. With three circular categories, all categories are adjacent to one another. Nominal categories are unordered, and thus none of the categories is adjacent to another category. It appears that from a mathematical perspective the two situations are quite similar.

4. Relationships Between Circular Kappas

In this section several relationships between the circular kappas are presented. One result involves all circular kappas (Lemma 2), while another result only applies to certain circular kappas (Lemma 4). Since all special cases coincide with $c = 3$ categories (Lemma 1), we assume from here on that $c \geq 4$. Lemma 2 shows that there exist precisely two orderings of the circular kappas.

Lemma 2. *Let $c \geq 4$ and $0 \leq u < v < 1$. We have $\kappa_u < \kappa_v$ if and only if*

$$\frac{\lambda_1}{\mu_1} > \frac{\lambda_2}{\mu_2}. \tag{13}$$

Proof. We first show that (14) is equivalent to (17). We have $\kappa_u < \kappa_v$ if and only if

$$\frac{O_u - E_u}{1 - E_u} < \frac{O_v - E_v}{1 - E_v}. \tag{14}$$

Since $1 - E_u$ and $1 - E_v$ are positive numbers, cross-multiplying the terms of (14) yields

$$O_u - E_u - O_u E_v < O_v - E_v - O_v E_u. \tag{15}$$

Adding $E_v - O_u + O_u E_u$ to both sides of (15) we obtain

$$(E_v - E_u)(1 - O_u) < (O_v - O_u)(1 - E_u). \tag{16}$$

Since $1 - E_v$ and $E_v - E_u$ are positive numbers, inequality (16) is equivalent to

$$\frac{1 - O_u}{1 - E_u} < \frac{O_v - O_u}{E_v - E_u}. \tag{17}$$

Next, using definitions (8) and (9), inequality (17) becomes

$$\frac{1 - \lambda_0 - u\lambda_1}{1 - \mu_0 - u\mu_1} < \frac{\lambda_1}{\mu_1}. \tag{18}$$

Inserting definitions (5c) and (6c) on the left-hand side of (18), we obtain

$$\frac{(1 - u)\lambda_1 + \lambda_2}{(1 - u)\mu_1 + \mu_2} < \frac{\lambda_1}{\mu_1}. \tag{19}$$

Cross-multiplying the terms of (19), followed by some algebra, we finally obtain inequality (13).



Lemma 2 shows that if inequality (13) holds all special cases of (10) are strictly ordered. In fact, the circular kappas can be ordered in precisely two ways. If (13) holds, Cohen’s kappa κ_0 has the smallest value and we have $\kappa_u < \kappa_v$ if $u < v$. Note that inequality (13) holds for the data in Table 1, since $\hat{\lambda}_2 = 0$ for these data. Furthermore, the inequality also holds for the data in Table 2, since

$$\frac{\lambda_1}{\mu_1} = \frac{0.27}{0.32} = 0.83 > 0.24 = \frac{0.12}{0.49} = \frac{\lambda_2}{\mu_2}. \tag{20}$$

Table 3 shows that the point estimates of the circular kappas are indeed ordered as predicted by Lemma 2.

The reverse ordering holds if the converse of condition (13) holds, that is, $\lambda_1/\mu_1 < \lambda_2/\mu_2$. In this case Cohen’s kappa κ_0 has the highest value and we have $\kappa_u > \kappa_v$ if $u < v$. All circular kappas coincide if $c = 2, 3$ (Lemma 1) and if (13) becomes an equality. This second ordering is less likely to occur, since it requires that there is more disagreement on categories that are not adjacent in the ordering than on categories that are adjacent.

The value of circular kappa κ_u is bounded by κ_0 and $\lim_{u \rightarrow 1} \kappa_u$. Lemma 3 presents an expression of this limit.

Lemma 3. *Let $c \geq 4$. It holds that*

$$\lim_{u \rightarrow 1} \kappa_u = 1 - \frac{\lambda_2}{\mu_2}. \tag{21}$$

Proof. Using (10) and definitions (5c) and (6c) we have

$$\lim_{u \rightarrow 1} \kappa_u = \frac{\lambda_0 + \lambda_1 - \mu_0 - \mu_1}{1 - \mu_0 - \mu_1} = \frac{1 - \lambda_2 - 1 + \mu_2}{1 - 1 + \mu_2} = \frac{\mu_2 - \lambda_2}{\mu_2}.$$



If inequality (13) holds, the minimum value of κ_u is obtained for $u = 0$, whereas the maximum value is given in (21). If $\lambda_2 = 0$ (see for example Table 1) the maximum value is 1. If the converse of condition (13) holds, that is, $\lambda_1/\mu_1 < \lambda_2/\mu_2$, then Cohen’s kappa κ_0 is the maximum value and (21) presents the minimum value.

Next, we consider a different type of relationship between the circular kappas. Lemma 4 below provides an interpretation for a specific class of circular kappas. It sometimes makes sense to combine two categories, for example, if two categories are not clearly defined or are easily confused. The disagreement between the categories can be removed by combining the

categories. Since the categories of a circular scale are ordered, it only makes sense to combine categories that are adjacent in the ordering. With c categories there are c adjacent pairs of categories, and thus c different ways to collapse an $c \times c$ agreement into an $(c - 1) \times (c - 1)$ agreement table.

It turns out that certain circular kappas are weighted averages of the values of Cohen’s kappa corresponding to the c collapsed tables that are obtained by combining two adjacent categories. This is proved in the following lemma.

Lemma 4. *Let $c \geq 4$. Furthermore, let $\kappa_0(i)$, $O_0(i)$ and $E_0(i)$ for $i \in \{1, 2, \dots, c - 1\}$ denote the values of, respectively, κ_0 , O_0 and E_0 of the $(c - 1) \times (c - 1)$ table that is obtained by combining categories i and $i + 1$. Moreover, let $\kappa_0(c)$, $O_0(c)$ and $E_0(c)$ denote the values of, respectively, κ_0 , O_0 and E_0 of the $(c - 1) \times (c - 1)$ table that is obtained by combining categories 1 and c . Then*

$$\kappa_{1/c} = \frac{\sum_{i=1}^c \kappa_0(i)(1 - E_0(i))}{\sum_{i=1}^c (1 - E_0(i))}. \tag{22}$$

Proof. Using (5a) we have

$$O_0(i) = \lambda_0 + \pi_{i,i+1} + \pi_{i+1,i}, \quad \text{for } i \in \{1, 2, \dots, c - 1\} \tag{23}$$

and

$$O_0(c) = \lambda_0 + \pi_{1c} + \pi_{c1}. \tag{24}$$

Then, using (5b), (23) and (24), we find that $\sum_{i=1}^c O_0(i)$ is equal to

$$\sum_{i=1}^c O_0(i) = c\lambda_0 + \lambda_1. \tag{25}$$

Using similar arguments, we find that $\sum_{i=1}^c E_0(i)$ is equal to

$$\sum_{i=1}^c E_0(i) = c\mu_0 + \mu_1. \tag{26}$$

The quantities $\kappa_0(i)$, $O_0(i)$ and $E_0(i)$ are related by the formula

$$\kappa_0(i) = \frac{O_0(i) - E_0(i)}{1 - E_0(i)}, \tag{27}$$

or equivalently, $\kappa_0(i)(1 - E_0(i)) = O_0(i) - E_0(i)$. Finally, using the latter identity, together with (25) and (26), we have

$$\begin{aligned} \frac{\sum_{i=1}^c \kappa_0(i)(1 - E_0(i))}{\sum_{i=1}^c (1 - E_0(i))} &= \frac{\sum_{i=1}^c (O_0(i) - E_0(i))}{c - \sum_{i=1}^c E_0(i)} \\ &= \frac{c\lambda_0 + \lambda_1 - c\mu_0 - \mu_1}{c - c\mu_0 - \mu_1} \\ &= \frac{\lambda_0 + \frac{1}{c}\lambda_1 - \mu_0 - \frac{1}{c}\mu_1}{1 - \mu_0 - \frac{1}{c}\mu_1} \\ &= \kappa_{1/c}. \end{aligned}$$

■

An interesting application of Lemma 4 occurs when we have $c = 4$ categories. Since all circular kappas coincide with $c = 3$ categories (Lemma 1), the circular kappa $\kappa_{0.25}$ can be interpreted as a weighted average of the four kappas of the collapsed 3×3 tables that are obtained by combining adjacent categories.

5. Dependence on the Number of Categories

In this section, a possible dependence of the circular kappas on the number of categories is studied. In Lemma 5, it is assumed that data can be described by the specific structure presented in (28). This data structure is perhaps not realistic, but it provides a theoretical result on the dependency of the circular kappas. Lemma 5 presents an example of a class of agreement tables for which all circular kappas are increasing in the number of categories c .

Lemma 5. *Let $c \geq 4$ and let $0 \leq a_0, a_1, a_2 \leq 1$ with $a_0 + a_1 + a_2 = 1$. Furthermore, let the entries of $\{\pi_{ij}\}$ be given by*

$$\pi_{ij} = \begin{cases} a_0/c, & \text{for } i = j; \\ a_1/2c, & \text{for } |i - j| = 1, \text{ or } |i - j| = c - 1; \\ a_2/c(c - 3), & \text{otherwise.} \end{cases} \quad (28)$$

Then κ_u is strictly increasing in c for all u .

Proof. Under the conditions of the lemma we have $\lambda_0 = a_0$, $\lambda_1 = a_1$ and $\lambda_2 = a_2$. Using the identity $a_0 + a_1 + a_2 = 1$ we also have

$$\pi_{i+} = \pi_{+i} = \frac{a_0}{c} + \frac{2a_1}{2c} + \frac{(c - 3)a_2}{c(c - 3)} = \frac{1}{c}, \quad \text{for } i \in \{1, 2, \dots, c\}. \quad (29)$$

Using identity (29) we have $\mu_0 = 1/c$ and $\mu_1 = 2/c$, and thus

$$\kappa_u = \frac{a_0 + ua_1 - \frac{1}{c} - \frac{2u}{c}}{1 - \frac{1}{c} - \frac{2u}{c}} = \frac{c(a_0 + ua_1) - 1 - 2u}{c - 1 - 2u}. \tag{30}$$

Using the right-hand side of (30) the first partial derivative with respect to c is given by

$$\begin{aligned} \frac{\partial}{\partial c} \kappa_u &= \frac{(a_0 + ua_1)(c - 1 - 2u) - c(a_0 + ua_1) + 1 + 2u}{(c - 1 - 2u)^2} \\ &= \frac{(1 + 2u)(1 - a_0 - ua_1)}{(c - 1 - 2u)^2}. \end{aligned} \tag{31}$$

Because $u < 1$, we have

$$a_0 + ua_1 < a_0 + a_1 \leq a_0 + a_1 + a_2 = 1, \tag{32}$$

and thus the inequality $a_0 + ua_1 < 1$. It follows from this latter inequality that (31) is strictly positive. Thus, under the conditions of the lemma κ_u is strictly increasing in c .

■

Lemma 5 shows that if we consider a series of agreement tables of a form (28) and keep the values of the total observed agreement λ_0 and the total disagreement on adjacent categories λ_1 fixed, then the values of the circular kappas increase with the size of the table. Using identity (30), we find that with a large number of categories the value of κ_u approaches

$$\lim_{c \rightarrow \infty} \kappa_u = \lim_{c \rightarrow \infty} \frac{c(\lambda_0 + u\lambda_1) - 1 - 2u}{c - 1 - 2u} = \lambda_0 + u\lambda_1. \tag{33}$$

6. Discussion

A family of kappa coefficients for assessing agreement between two circular classifications with identical categories was presented. If the categories form a circular scale, the categories exhibit a certain periodicity, and the designation of high and low is arbitrary. The standard weighted kappas used with linear scales, that is, linear and quadratic kappa (Vanbelle 2015; Warrens 2013, 2015), are not appropriate for analyzing agreement between circular classifications.

The following properties of the circular kappas were formally proved. The circular kappas all coincide if the agreement table has two or three categories (Lemma 1). Furthermore, the values of the circular kappas can be strictly ordered in precisely two ways (Lemma 2). Moreover, certain circular kappas can be interpreted as weighted averages of the Cohen’s kappas of

all collapsed tables that are obtained by combining two adjacent categories (Lemma 4). Finally, for a particular type of agreement table it was shown that the values of the circular kappas increase with the number of categories (Lemma 5).

The values of the circular kappas can be strictly ordered in two different ways, but one ordering is more likely to occur in practice than the other. In the likely ordering (see Tables 1 and 2), Cohen's kappa produces the smallest value, and the values of the circular kappas increase as more weight is assigned to the total disagreement on adjacent categories. The strict ordering suggests that the circular kappas are measuring the same concept, but to a different extent. This in turn suggests that we might as well use the most well-known kappa coefficient in this family, which is Cohen's kappa. Using Cohen's kappa has several advantages. First of all, if we use Cohen's kappa it is not necessary to specify a positive value of the parameter u , which is arbitrary to some extent. Secondly, Cohen's kappa has been applied in thousands of applications and many of its properties are well understood (Zhou and Yang 2014; Warrens 2008, 2010b, 2013).

Various authors have presented target values for evaluating the values of kappa coefficients (Landis and Koch 1977; Altman 1991). For example, a value of 0.80 for Cohen's kappa generally indicates good or excellent agreement. There is general consensus in the literature that uncritical application of such magnitude guidelines leads to practically questionable decisions. Since the circular kappas tend to measure the same thing, and since circular kappas that give a large weight to the total disagreement on adjacent categories appear to produce values that are substantially higher than the values of the circular kappas that give a small weight to the total disagreement on adjacent categories, the same guidelines cannot be applied to all circular kappas. If one accepts the use of magnitude guidelines, it seems reasonable to use stricter criteria for circular kappas that tend to produce high values.

If one is interested in using a circular kappa other than Cohen's kappa, then Lemma 4 provides an interesting case when we have $c = 4$ categories. Since all circular kappas coincide with $c = 3$ categories (Lemma 1), the circular kappa $\kappa_{0.25}$ can be interpreted as a weighted average of the four kappas of the collapsed 3×3 tables that are obtained by combining adjacent categories. Since $\kappa_{0.25}$ is an average, its value lies between the minimum and maximum value of the four kappas of the collapsed 3×3 tables. Furthermore, because these four kappas usually (with real life data) have distinct values, it follows that there exist two categories such that, when combined, the value of $\kappa_{0.25}$ is increased. Moreover, there exist two categories such that, when combined, the value of $\kappa_{0.25}$ is decreased. This minor existence result does not tell us which categories these are, just that they exist.

References

- ALTMAN, D.G. (1991), *Practical Statistics for Medical Research*, London: Chapman & Hall.
- BERENS, P. (2009), "Circstat: A MATLAB Toolbox for Circular Statistics", *Journal of Statistical Software*, 31, 1–21.
- BROWN, M.W. (1992), "Circumplex Models for Correlation Matrices", *Psychometrika*, 57, 470–479.
- FLEISS, J.L., COHEN, J., and EVERITT, B.S. (1969), "Large Sample Standard Errors of Kappa and Weighted Kappa", *Psychological Bulletin*, 72, 323–327.
- GWET, K.L. (2012), *Handbook of Inter-Rater Reliability* (3rd ed.), Gaithersburg MD: Advanced Analytics LLC.
- LANDIS, J.R., and KOCH, G.G. (1977), "The Measurement of Observer Agreement for Categorical Data", *Biometrics*, 33, 159–174.
- POSNER, J., RUSSELL, J.A., and PETERSON, B.S. (2005), "The Circumplex Model of Affect: An Integrative Approach to Affective Neuroscience, Cognitive Development, and Psychopathology", *Developmental Psychopathology*, 17, 715–734.
- RUEDA, C., FERNÁNDEZ, M.A., and PEDDADA, S.D. (2009), "Estimation of Parameters Subject to Order Restrictions on a Circle With Application to Estimation of Phase Angles of Cell Cycle Genes", *Journal of the American Statistical Association*, 104, 338–347.
- RUSSELL, J.A. (1980), "A Circumplex Model of Affect", *Journal of Personality and Social Psychology*, 39, 1161–1178.
- VANBELLE, S. (2015), "A New Interpretation of the Weighted Kappa Coefficients", *Psychometrika*, in press.
- VANBELLE, S., MUTSVARI, T., DECLERCK, D., and LESAFFRE, E. (2012), "Hierarchical Modeling of Agreement", *Statistics in Medicine*, 31, 3667–3680.
- WARRENS, M.J. (2008), "On the Equivalence of Cohen's Kappa and the Hubert-Arabie Adjusted Rand Index", *Journal of Classification*, 25, 177–183.
- WARRENS, M.J. (2010a), "A Kraemer-Type Rescaling that Transforms the Odds Ratio Into the Weighted Kappa Coefficient", *Psychometrika*, 75, 328–330.
- WARRENS, M.J. (2010b), "Inequalities Between Multi-rater Kappas", *Advances in Data Analysis and Classification*, 4, 271–286.
- WARRENS, M.J. (2011a), "Cohen's Kappa is a Weighted Average", *Statistical Methodology*, 8, 473–484.
- WARRENS, M.J. (2011b), "Weighted Kappa is Higher Than Cohen's Kappa for Tridiagonal Agreement Tables", *Statistical Methodology*, 8, 268–272.
- WARRENS, M.J. (2012), "Some Paradoxical Results for the Quadratically Weighted Kappa", *Psychometrika*, 77, 315–323.
- WARRENS, M.J. (2013), "Cohen's Weighted Kappa With Additive Weights", *Advances in Data Analysis and Classification*, 7, 41–55.
- WARRENS, M.J. (2014), "Corrected Zegers-ten Berge Coefficients Are Special Cases of Cohen's Weighted Kappa", *Journal of Classification*, 31, 179–193.
- WARRENS, M.J. (2015), "Additive Kappa Can be Increased by Combining Adjacent Categories", *International Mathematical Forum*, 10, 323–328.
- WATSON, D., and TELLEGEN, A. (1985), "Toward a Consensual Structure of Mood", *Psychological Bulletin*, 98, 219–235.

- WATSON, D., WIESE, D., VAIDYA, J., and TELLEGEN, A. (1999), “The Two General Activation Systems of Affect: Structural Findings, Evolutionary Considerations, and Psychobiological Evidence”, *Journal of Personality and Social Psychology*, 76, 820–838.
- YANG, Z., and ZHOU, M. (2014), “Kappa Statistic for Clustered Matched-pair Data”, *Statistics in Medicine*, 33, 2612–2633.
- YANG, Z., and ZHOU, M. (2015), “Weighted Kappa Statistic for Clustered Matched-Pair Ordinal Data”, *Computational Statistics and Data Analysis*, 82, 1–18.

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.