Cohen's quadratically weighted kappa is higher than linearly weighted kappa for tridiagonal agreement tables
Warrens, M.J.

Citation

Version: Not Applicable (or Unknown)
License: Leiden University Non-exclusive license
Downloaded from: https://hdl.handle.net/1887/18381

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

http://dx.doi.org/10.1016/j.stamet.2011.08.006

Author. Matthijs J. Warrens
Institute of Psychology
Unit Methodology and Statistics
Leiden University
P.O. Box 9555, 2300 RB Leiden
The Netherlands
E-mail: warrens@fsw.leidenuniv.nl
Cohen’s quadratically weighted kappa is higher than linearly weighted kappa for tridiagonal agreement tables

Matthijs J. Warrens, Leiden University

Abstract: Cohen’s weighted kappa is a popular descriptive statistic for measuring the agreement between two raters on an ordinal scale. Popular weights for weighted kappa are the linear weights and the quadratic weights. It has been frequently observed in the literature that the value of the quadratically weighted kappa is higher than the value of the linearly weighted kappa. In this paper this phenomenon is proved for tridiagonal agreement tables. A square table is tridiagonal if it has nonzero elements only on the main diagonal and on the two diagonals directly adjacent to the main diagonal.

Key words: Cohen’s kappa; Ordinal agreement; Linear weights; Quadratic weights.
1 Introduction

The kappa coefficient (denoted by \( \kappa \)) is a widely used descriptive statistic for summarizing two nominal variables with identical categories [2, 5, 19, 20, 21, 22, 25, 26]. Cohen’s \( \kappa \) was originally proposed as a measure of agreement between two raters (observers) who rate each of the same sample of objects (individuals, observations) on a nominal scale with \( n \in \mathbb{N}_{\geq 2} \) mutually exclusive categories. The \( \kappa \) statistic has been applied to numerous agreement tables encountered in psychology, educational measurement and epidemiology. The value of \( \kappa \) is 1 when perfect agreement between the two raters occurs, 0 when agreement is equal to that expected under independence, and negative when agreement is less than that expected by chance. The popularity of \( \kappa \) has led to the development of many extensions [1, 11, 23].

A popular generalization of Cohen’s \( \kappa \) is the weighted kappa coefficient (denoted by \( \kappa_w \)) which was proposed for situations where the disagreements between the raters are not all equally important [6, 9, 10, 13, 16, 25]. For example, when categories are ordered, the seriousness of a disagreement depends on the difference between the ratings. Cohen’s \( \kappa_w \) allows the use of weights to describe the closeness of agreement between categories. Popular weights are the so-called linear weights [4, 12, 16] and the quadratic weights [9, 13]. In this paper the linearly weighted kappa will be denoted by \( \kappa_1 \), whereas the quadratically weighted kappa will be denoted by \( \kappa_2 \).

A frequent criticism against the use of \( \kappa_w \) is that the weights are arbitrarily defined [16]. In support of \( \kappa_2 \) it turns out that \( \kappa_2 \) is equivalent to the product-moment correlation coefficient under specific conditions [6]. In addition, \( \kappa_2 \) may be interpreted as an intraclass correlation coefficient [9, 13]. In support of \( \kappa_1 \) it turns out that the components of \( \kappa_1 \) corresponding to an \( n \times n \) agreement table can be obtained from the \( n - 1 \) distinct collapsed 2 \( \times \) 2 tables that are obtained by combining adjacent categories [16].

It has been frequently observed in the literature that the value of \( \kappa_2 \) is higher than the value of \( \kappa_1 \). For example, consider the data in Table 1 taken from a study in [15]. In this study 100 patients were rated by two randomly allocated observers on their degree of handicap. For these data we have \( \kappa_1 = 0.780 < 0.907 = \kappa_2 \). A value of 1 would indicate perfect agreement between the observers. The value of \( \kappa_2 \) does not always exceeds the value of \( \kappa_1 \). It turns out however that the inequality holds for a special kind of agreement table. In this paper we prove that \( \kappa_2 > \kappa_1 \) when the agreement table is tridiagonal. A tridiagonal table is a square matrix that has nonzero elements only on the main diagonal and on the two diagonals directly adjacent to the main diagonal [25]. Note that Table 1 is almost tridiagonal. Agreement tables that are tridiagonal or approximately tridiagonal are frequently observed in
Table 1: Ratings of 100 patients by pairs of observers on the degree of disability on a 6-category scale [15].

<table>
<thead>
<tr>
<th>Observer 2</th>
<th>Observer 1</th>
<th>Row totals</th>
</tr>
</thead>
<tbody>
<tr>
<td>0 = No symptoms</td>
<td>5 6 1 3 2 1</td>
<td>5 10 21 16 22 25 100</td>
</tr>
<tr>
<td>1 = Not significant disability</td>
<td>2</td>
<td>8</td>
</tr>
<tr>
<td>2 = Slight disability</td>
<td>1 4 13 5 2</td>
<td>25</td>
</tr>
<tr>
<td>3 = Moderate disability</td>
<td>2 8 1</td>
<td>11</td>
</tr>
<tr>
<td>4 = Moderately severe dis.</td>
<td>8 24 32</td>
<td></td>
</tr>
<tr>
<td>5 = Severe disability</td>
<td>6 10 21 16 22 25 100</td>
<td></td>
</tr>
</tbody>
</table>

applications with ordered categories [3, 7, 8, 14].

The paper is organized as follows. In the next section we define a particular case of $\kappa_w$, denoted by $\kappa_m$, of which $\kappa_1$ and $\kappa_2$ are special cases. The main result, a conditional inequality between $\kappa_m$ and $\kappa_\ell$ for $m > \ell \geq 1$, is presented in Section 3. The result depicted in the title of this paper is an immediate consequence of the main result.

2 Cohen’s weighted kappa

Suppose that two observers each distribute the same set of $k \in \mathbb{N}_{\geq 1}$ objects (individuals) among a set of $n \in \mathbb{N}_{\geq 2}$ mutually exclusive categories that are defined in advance. Let $F = (f_{ij})$ with $i, j \in \{1, 2, \ldots, n\}$ be the agreement table with the ratings of the observers, where $f_{ij}$ indicates the number of objects placed in category $i$ by the first observer and in category $j$ by the second observer. We assume that the categories of observers are in the same order so that the diagonal elements $f_{ii}$ reflect the number of objects put in the same categories by the observers. For notational convenience we work with the table of proportions $P = (p_{ij})$ with relative frequencies $p_{ij} = f_{ij}/k$.

Row and column totals

$$p_i = \sum_{j=1}^{n} p_{ij} \quad \text{and} \quad q_i = \sum_{j=1}^{n} p_{ji}$$
are the marginal totals of $P$. The weighted kappa statistic can be defined as

$$\kappa_w = \frac{O_w - E_w}{1 - E_w}$$

(1)

where

$$O_w = \sum_{i,j=1}^{n} w_{ij} p_{ij} \quad \text{and} \quad E_w = \sum_{i,j=1}^{n} w_{ij} p_i q_j.$$ 

For the weights $w_{ij}$ we require $w_{ij} \in [0, 1]$ and $w_{ii} = 1$ for $i, j \in \{1, 2, \ldots, n\}$. In (1) we assume that $E_w < 1$ to avoid the indeterminate case $E_w = 1$. If we use $w_{ij} = 1$ if $i = j$ and $w_{ij} = 0$ if $i \neq j$ for $i, j \in \{1, 2, \ldots, n\}$, $\kappa_w$ is equal to Cohen’s unweighted $\kappa$.

Examples of weights for $\kappa_w$ that have been proposed in the literature, are the linear weights [4, 12, 16, 24] given by

$$w_{ij}^{(1)} = 1 - \frac{|i - j|}{n - 1}$$

(2)

and the quadratic weights [9, 13] given by

$$w_{ij}^{(2)} = 1 - \left( \frac{i - j}{n - 1} \right)^2.$$ 

(3)

Let $m \in \mathbb{R}_{\geq 1}$. The weights in (2) and (3) are special cases of the family of weights given by

$$w_{ij}^{(m)} = 1 - \left( \frac{|i - j|}{n - 1} \right)^m \quad \text{for} \quad m \geq 1.$$ 

In this paper we are particularly interested in the special case of $\kappa_w$ given by

$$\kappa_m = \frac{O_m - E_m}{1 - E_m}$$

(4)

where

$$O_m = \sum_{i,j=1}^{n} w_{ij}^{(m)} p_{ij} \quad \text{and} \quad E_m = \sum_{i,j=1}^{n} w_{ij}^{(m)} p_i q_j.$$ 

Special cases of $\kappa_m$ are the linearly weighted kappa $\kappa_1$ and the quadratically weighted kappa $\kappa_2$. We have $\kappa = \kappa_m$ in the case of $n = 2$ categories [17, 18, 19] and if $O_m = 1$. For the data in Table 1 we have $O_1 = 0.924$, $E_1 = 0.655$ and $\kappa_1 = 0.780$, and $O_2 = 0.982$, $E_2 = 0.811$ and $\kappa_2 = 0.907$. 

5
3 A conditional inequality

Theorem 1 shows that, for $m > \ell \geq 1$, $\kappa_m > \kappa_\ell$ if $P$ is tridiagonal. The latter concept is captured in the following definition.

Definition. A square agreement table $P$ is called tridiagonal if the only nonzero elements of $P$ are the $p_{ii}$ for $i \in \{1, 2, \ldots, n\}$, and the $p_{i,i+1}$ and $p_{i+1,i}$ for $i \in \{1, 2, \ldots, n-1\}$.

Theorem 1. Let $n \geq 3$ and let $m > \ell \geq 1$. Furthermore, suppose that $P$ is tridiagonal and that not all the $p_{i,i+1}$ and $p_{i+1,i}$ are 0. Then $\kappa_m > \kappa_\ell$.

Proof: We first show that (5) is equivalent to (9). Since $1 - E_\ell$ and $1 - E_m$ are positive numbers, we have $\kappa_m > \kappa_\ell$ if and only if

$$
\frac{O_m - E_m}{1 - E_m} > \frac{O_\ell - E_\ell}{1 - E_\ell}
$$

(5)

$$
\text{⇓}
$$

$$
(O_m - E_m)(1 - E_\ell) > (O_\ell - E_\ell)(1 - E_m)
$$

(6)

Subtracting $O_\ell + E_\ell E_m$ from and adding $E_m + O_\ell E_\ell$ to both sides of (6), we obtain

$$
(O_m - O_\ell)(1 - E_\ell) > (E_m - E_\ell)(1 - O_\ell).
$$

(7)

Let $w^{(\ell)}$ and $w^{(m)}$ denote the weights of $p_{i,i+1}$ and $p_{i+1,i}$ respectively for $\kappa_\ell$ and $\kappa_m$. We have

$$
w^{(m)} - w^{(\ell)} = \frac{1}{(n-1)\ell} - \frac{1}{(n-1)m}.
$$

(8)

Since $m > \ell \geq 1$ it follows from (8) that $w^{(m)} - w^{(\ell)} > 0$. Furthermore, since not all the $p_{i,i+1}$ and $p_{i+1,i}$ are 0, there is an element on one of the diagonals adjacent to the main diagonal for which the weights satisfy $w^{(m)} - w^{(\ell)} > 0$. Hence $E_m - E_\ell > 0$, and inequality (7) is equivalent to the inequality

$$
\frac{O_m - O_\ell}{E_m - E_\ell} > \frac{1 - O_\ell}{1 - E_\ell}.
$$

(9)

Next, if $P$ is tridiagonal inequality (9) becomes

$$
\frac{(w^{(m)} - w^{(\ell)}) \sum_{i=1}^{n-1} (p_{i,i+1} + p_{i+1,i})}{\sum_{i,j=1}^{n} (w^{(m)}_{ij} - w^{(\ell)}_{ij})p_{ij}q_{j}} > \frac{(1 - w^{(\ell)}) \sum_{i=1}^{n-1} (p_{i,i+1} + p_{i+1,i})}{\sum_{i,j=1}^{n} (1 - w^{(\ell)}_{ij})p_{ij}q_{j}}.
$$

(10)
Since $\sum_{i=1}^{n-1}(p_{i,i+1} + p_{i+1,i}) > 0$ (not all the $p_{i,i+1}$ and $p_{i+1,i}$ are 0), (10) is equal to the inequality

$$\sum_{i,j=1}^{n} \left[ (w^{(m)}_{ij} - w^{(\ell)}_{ij})(1 - w^{(\ell)}_{ij}) - (1 - w^{(m)}_{ij})(w^{(m)}_{ij} - w^{(\ell)}_{ij}) \right] p_{ij} q_{ij} > 0. \tag{11}$$

For $|i-j| = 0$ we have $w^{(\ell)}_{ij} = w^{(m)}_{ij} = 1$, whereas for $|i-j| = 1$ we have $w^{(\ell)}_{ij} = w^{(\ell)}$ and $w^{(m)}_{ij} = w^{(m)}$. In both cases we have $(w^{(m)} - w^{(\ell)})(1 - w^{(\ell)}_{ij}) = (1 - w^{(\ell)})(w^{(m)}_{ij} - w^{(\ell)}_{ij})$. Hence, inequality (11) holds if

$$(w^{(m)} - w^{(\ell)})(1 - w^{(\ell)}_{ij}) - (1 - w^{(m)}_{ij})(w^{(m)}_{ij} - w^{(\ell)}_{ij}) > 0 \tag{12}$$

for $|i-j| \geq 2$.

We have

$$1 - w^{(\ell)}_{ij} = \left(\frac{|i-j|}{n-1}\right)^{\ell} \tag{13a}$$
$$1 - w^{(\ell)} = \frac{1}{(n-1)^{\ell}} \tag{13b}$$
$$w^{(m)}_{ij} - w^{(\ell)}_{ij} = \left(\frac{|i-j|}{n-1}\right)^{\ell} - \left(\frac{|i-j|}{n-1}\right)^{m}. \tag{13c}$$

Using the identities in (8) and (13), inequality (12) is equal to

$$\left[ \left(\frac{1}{n-1}\right)^{\ell} - \left(\frac{1}{n-1}\right)^{m} \right] \left(\frac{|i-j|}{n-1}\right)^{\ell} > \left(\frac{1}{n-1}\right)^{\ell} \left[ \left(\frac{|i-j|}{n-1}\right)^{\ell} - \left(\frac{|i-j|}{n-1}\right)^{m} \right]$$
$$\Updownarrow$$
$$\left(\frac{1}{n-1}\right)^{\ell} \left(\frac{|i-j|}{n-1}\right)^{m} > \left(\frac{1}{n-1}\right)^{m} \left(\frac{|i-j|}{n-1}\right)^{\ell}$$
$$\Updownarrow$$
$$\left(\frac{|i-j|}{n-1}\right)^{m-\ell} > \left(\frac{1}{n-1}\right)^{m-\ell}. \tag{14}$$

Inequality (14) and thus inequality (12) hold for $|i-j| \geq 2$, and hence inequality (11) is valid. This completes the proof. $\square$

Recall that $\kappa$ denotes Cohen’s unweighted kappa. Since $\kappa_m$ satisfies the conditions of the theorem in [25] we have the following result.
Corollary 1. Let \( n \geq 3 \). Furthermore, suppose that \( P \) is tridiagonal and that not all the \( p_{i,i+1} \) and \( p_{i+1,i} \) are 0. Then \( \kappa_m > \kappa \).

Thus, the value of Cohen’s \( \kappa \) never exceeds the value of \( \kappa_m \) if the agreement table is tridiagonal.

The result depicted in the title of this paper is an immediate consequence of Theorem 1.

Corollary 2. Let \( n \geq 3 \). Furthermore, suppose that \( P \) is tridiagonal and that not all the \( p_{i,i+1} \) and \( p_{i+1,i} \) are 0. Then \( \kappa_2 > \kappa_1 > \kappa \).

References


