Cohen's quadratically weighted kappa is higher than linearly weighted kappa for tridiagonal agreement tables

Matthijs J. Warrens *

Department of Methodology and Statistics, Tilburg University, P.O. Box 90153, 5000 LE, Tilburg, Netherlands

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.

© 2011 Elsevier B.V. All rights reserved.

1. Introduction

The kappa coefficient (denoted by κ) is a widely used descriptive statistic for summarizing two nominal variables with identical categories [2,5,19,21,20,22,25,26]. Cohen's κ 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 κ statistic has been applied to numerous agreement tables encountered in psychology, educational measurement and epidemiology. The value of κ 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 κ has led to the development of many extensions [1,11,23].

A popular generalization of Cohen's κ is the weighted kappa coefficient (denoted by κ_w) which was proposed for situations where the disagreements between the raters are not all equally important [6,9,10,13,15,25]. For example, when categories are ordered, the seriousness of a disagreement

---

* Tel.: +31 13 4663270; fax: +31 13 4663002.
  E-mail address: m.j.warrens@uvt.nl.

Table 1
Ratings of 100 patients by pairs of observers on the degree of disability on a 6-category scale [16].

Observer 2	Observer 1						Row totals
	0	1	2	3	4	5	
0 = No symptoms	5						5
1 = Not significant disability		6	2				8
2 = Slight disability	1	4	13	5	2		25
3 = Moderate disability			6	9	4		19
4 = Moderately severe dis.				2	8	1	11
5 = Severe disability					8	24	32
Column totals	6	10	21	16	22	25	100

depends on the difference between the ratings. Cohen's κ_w allows the use of weights to describe the closeness of agreement between categories. Popular weights are the so-called linear weights [4,12,15] and the quadratic weights [9,13]. In this paper, the linearly weighted kappa will be denoted by κ_1, whereas the quadratically weighted kappa will be denoted by κ_2.

A frequent criticism against the use of κ_w is that the weights are arbitrarily defined [15]. In support of κ_2 it turns out that κ_2 is equivalent to the product-moment correlation coefficient under specific conditions [6]. In addition, κ_2 may be interpreted as an intraclass correlation coefficient [9,13]. In support of κ_1 it turns out that the components of κ_1 corresponding to an $n \times n$ agreement table can be obtained from the $n - 1$ distinct collapsed 2×2 tables that are obtained by combining adjacent categories [15].

It has been frequently observed in the literature that the value of κ_2 is higher than the value of κ_1. For example, consider the data in Table 1 taken from a study in [16]. 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 κ_2 does not always exceeds the value of κ_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 applications with ordered categories [3,7,8,14].

The paper is organized as follows. In the next section, we define a particular case of κ_w, denoted by κ_m, of which κ_1 and κ_2 are special cases. The main result, a conditional inequality between κ_m and κ_ℓ 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

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

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 $i, j \in \{1, 2, \ldots, n\}$, κ_w is equal to Cohen's unweighted κ.

Examples of weights for κ_w that have been proposed in the literature, are the linear weights [4,12,15,24] given by

$$w_{ij}^{(1)} = 1 - \frac{|i - j|}{n - 1} \tag{2}$$

and the quadratic weights [9,13] given by

$$w_{ij}^{(2)} = 1 - \left(\frac{i - j}{n - 1} \right)^2. \tag{3}$$

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

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

In this paper, we are particularly interested in the special case of κ_w given by

$$\kappa_m = \frac{O_m - E_m}{1 - E_m} \tag{4}$$

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 κ_m are the linearly weighted kappa κ_1 and the quadratically weighted kappa κ_2. We have $\kappa = \kappa_m$ in the case of $n = 2$ categories [17–19] 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 $\kappa_2 = 0.907$.

3. A conditional inequality

The theorem below 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. *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

$$\frac{O_m - E_m}{1 - E_m} > \frac{O_\ell - E_\ell}{1 - E_\ell} \tag{5}$$

$$\Updownarrow$$

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

$$\Updownarrow$$

$$O_m - E_m - O_m E_\ell + E_m E_\ell > O_\ell - E_\ell - O_\ell E_m + E_\ell E_m. \tag{6}$$

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

$$(O_m - O_\ell)(1 - E_\ell) > (E_m - E_\ell)(1 - O_\ell). \tag{7}$$

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

$$w^{(m)} - w^{(\ell)} = \frac{1}{(n-1)^\ell} - \frac{1}{(n-1)^m}. \tag{8}$$

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

$$\frac{O_m - O_\ell}{E_m - E_\ell} > \frac{1 - O_\ell}{1 - E_\ell}. \tag{9}$$

Next, if P is tridiagonal inequality (9) becomes

$$\frac{(w^{(m)} - w^{(\ell)}) \sum_{i=1}^{n-1} (p_{i,i+1} + p_{i+1,i})}{\sum_{i,j=1}^{n} (w_{ij}^{(m)} - w_{ij}^{(\ell)}) p_i q_j} > \frac{(1 - w^{(\ell)}) \sum_{i=1}^{n-1} (p_{i,i+1} + p_{i+1,i})}{\sum_{i,j=1}^{n} (1 - w_{ij}^{(\ell)}) p_i q_j}. \tag{10}$$

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

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

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

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

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

We have

$$1 - w_{ij}^{(\ell)} = \left(\frac{|i - j|}{n - 1} \right)^\ell \tag{13a}$$

$$1 - w^{(\ell)} = \frac{1}{(n-1)^\ell} \tag{13b}$$

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

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 κ denotes Cohen's unweighted kappa. Since κ_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 κ never exceeds the value of κ_m if the agreement table is tridiagonal. The result depicted in the title of this paper is an immediate consequence of the theorem above.

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

[1] M. Banerjee, M. Capozzoli, L. McSweeney, D. Sinha, Beyond kappa: a review of interrater agreement measures, The Canadian Journal of Statistics 27 (1999) 3–23.
[2] R.L. Brennan, D.J. Prediger, Coefficient kappa: some uses, misuses, and alternatives, Educational and Psychological Measurement 41 (1981) 687–699.
[3] J.-M. Cai, T.S. Hatsukami, M.S. Ferguson, R. Small, N.L. Polissar, C. Yuan, Classification of human carotid atherosclerotic lesions with in vivo multicontrast magnetic resonance imaging, Circulation 106 (2002) 1368–1373.
[4] D. Cicchetti, T. Allison, A new procedure for assessing reliability of scoring EEG sleep recordings, The American Journal of EEG Technology 11 (1971) 101–109.
[5] J. Cohen, A coefficient of agreement for nominal scales, Educational and Psychological Measurement 20 (1960) 37–46.
[6] J. Cohen, Weighted kappa: nominal scale agreement with provision for scaled disagreement or partial credit, Psychological Bulletin 70 (1968) 213–220.
[7] M.S. Dirksen, J.J. Bax, A. De Roos, J.W. Jukema, R.J. Van der Geest, K. Geleijns, E. Boersma, E.E. Van der Wall, H.J. Lamb, Usefulness of dynamic multislice computed tomography of left ventricular function in unstable angina pectoris and comparison with echocardiography, The American Journal of Cardiology 90 (2002) 1157–1160.
[8] W.W. Eaton, K. Neufeld, L.-S. Chen, G. Cai, A comparison of self-report and clinical diagnostic interviews for depression, Archives of General Psychiatry 57 (2000) 217–222.
[9] J.L. Fleiss, J. Cohen, The equivalence of weighted kappa and the intraclass correlation coefficient as measures of reliability, Educational and Psychological Measurement 33 (1973) 613–619.
[10] J.L. Fleiss, J. Cohen, B.S. Everitt, Large sample standard errors of kappa and weighted kappa, Psychological Bulletin 72 (1969) 323–327.
[11] H.C. Kraemer, V.S. Periyakoil, A. Noda, Tutorial in biostatistics: kappa coefficients in medical research, Statistics in Medicine 21 (2004) 2109–2129.
[12] P.W. Mielke, K.J. Berry, A note on Cohen's weighted kappa coefficient of agreement with linear weights, Statistical Methodology 6 (2009) 439–446.
[13] C. Schuster, A note on the interpretation of weighted kappa and its relations to other rater agreement statistics for metric scales, Educational and Psychological Measurement 64 (2004) 243–253.
[14] J.M. Seddon, C.R. Sahagian, R.J. Glynn, R.D. Sperduto, E.S. Gragoudas, Evaluation of an iris color classification system and the eye disorders case-control study group, Investigative Ophthalmology & Visual Science 31 (1990) 1592–1598.
[15] S. Vanbelle, A. Albert, A note on the linearly weighted kappa coefficient for ordinal scales, Statistical Methodology 6 (2009) 157–163.
[16] J.C. Van Swieten, P.J. Koudstaal, M.C. Visser, H.J.A. Schouten, J. Van Gijn, Interobserver agreement for the assessment of handicap in stroke patients, Stroke 19 (1987) 604–607.
[17] M.J. Warrens, On association coefficients for 2×2 tables and properties that do not depend on the marginal distributions, Psychometrika 73 (2008) 777–789.
[18] M.J. Warrens, On similarity coefficients for 2×2 tables and correction for chance, Psychometrika 73 (2008) 487–502.
[19] M.J. Warrens, On the equivalence of Cohen's kappa and the Hubert–Arabie adjusted rand index, Journal of Classification 25 (2008) 177–183.
[20] M.J. Warrens, A formal proof of a paradox associated with Cohen's kappa, Journal of Classification 27 (2010) 322–332.
[21] M.J. Warrens, Cohen's kappa can always be increased and decreased by combining categories, Statistical Methodology 7 (2010) 673–677.
[22] M.J. Warrens, Inequalities between kappa and kappa-like statistics for $k \times k$ tables, Psychometrika 75 (2010) 176–185.
[23] M.J. Warrens, Inequalities between multi-rater kappas, Advances in Data Analysis and Classification 4 (2010) 271–286.
[24] M.J. Warrens, Cohen's linearly weighted kappa is a weighted average of 2x2 kappas, Psychometrika 76 (2011) 471–486.
[25] M.J. Warrens, Weighted kappa is higher than Cohen's kappa for tridiagonal agreement tables, Statistical Methodology 8 (2011) 268–272.
[26] R. Zwick, Another look at interrater agreement, Psychological Bulletin 103 (1988) 374–378.