Hindawi Publishing Corporation
Advances in Statistics
Volume 2014, Article ID 853090, 5 pages
http://dx.doi.org/10.1155/2014/853090

Research Article

On Agreement Tables with Constant Kappa Values

Matthijs J. Warrens

Institute of Psychology, Unit Methodology and Statistics, Leiden University, P.O. Box 9555, 2300 RB Leiden, The Netherlands

Correspondence should be addressed to Matthijs J. Warrens; warrens@fsw.leidenuniv.nl

Received 8 July 2014; Accepted 14 August 2014; Published 24 August 2014

Academic Editor: Chin-Shang Li

Kappa coefficients are standard tools for summarizing the information in cross-classifications of two categorical variables with identical categories, here called agreement tables. When two categories are combined the kappa value usually either increases or decreases. There is a class of agreement tables for which the value of Cohen's kappa remains constant when two categories are combined. It is shown that for this class of tables all special cases of symmetric kappa coincide and that the value of symmetric kappa is not affected by any partitioning of the categories.

1. Introduction

In behavioral and biomedical science researchers are often interested in measuring the intensity of a behavior or a disease. Examples are psychologists that assess how anxious a speech-anxious subject appears while giving a talk, pathologists that rate the severity of lesions from scans, or competing diagnostic devices that classify the extent of a disease in patients into categories. These phenomena are typically classified using a categorical rating system, for example, with categories (A) slight, (B) moderate, and (C) extreme. Because ratings usually entail a certain degree of subjective judgment, researchers frequently want to assess the reliability of the categorical rating system that is used. One way to do this is to assign two observers to rate independently the same set of subjects. The reliability of the rating system can then be assessed by analyzing the agreement between the observers. High agreement between the ratings can be seen as a good indication of consensus in the diagnosis and interchangeability of the ratings of the observers.

Various statistical methodologies have been developed for analyzing agreement of a categorical rating system [1, 2]. For instance, loglinear models can be used for studying the patterns of agreement and sources of disagreement [3, 4]. However, in practice researchers often want to express the agreement between the raters in a single number. In this context, standard tools for summarizing agreement between observers are coefficients Cohen's kappa in the case of nominal categories [5–7] and weighted kappa in the case of ordinal categories [8–11]. With ordinal categories one may expect more disagreement or confusion on adjacent categories than on categories that are further apart. Weighted kappa allows the user to specify weights to describe the closeness between categories [12]. Both Cohen's kappa and weighted kappa are corrected for agreement due to chance. The coefficients were originally proposed in the context of agreement studies, but nowadays they are used for summarizing all kinds of cross-classifications of two variables with the same categories [11, 12].

The number of categories used in various rating systems usually varies from the minimum number of two to five in many practical applications. It is sometimes desirable to combine some of the categories [7]. For example, when two categories are easily confused, combining the categories usually improves the reliability of the rating system [13]. By collapsing categories the number of categories of the rating system is reduced. If there is a lot of disagreement between two categories, we expect the kappa value to increase if we combine the categories. This is usually the case. However, Schouten [13] showed that there is a class of agreement tables for which the value of Cohen's kappa remains constant when categories are merged. This is not what one expects from an agreement coefficient like Cohen's kappa. The question, then, arises: do other (weighted) kappa coefficients exhibit

Table 1: Two hypothetical 3 × 3 agreement tables.

First observer	Second observer							
	A	B	C	Total	A	B	C	Total
A	22	2	0	**24**	16	4	0	**20**
B	4	10	0	**14**	0	2	1	**3**
C	4	2	6	**12**	4	0	2	**6**
Total	**30**	**14**	**6**	**50**	**20**	**6**	**3**	**29**

Table 2: Two weighting schemes for four categories A, B, C, and D.

	Identity				Quadratic			
	A	B	C	D	A	B	C	D
A	0	1	1	1	0	1	4	9
B	1	0	1	1	1	0	1	4
C	1	1	0	1	4	1	0	1
D	1	1	1	0	9	4	1	0

the same property for these tables? If the answer is negative, it would make sense to replace Cohen's kappa by a weighted kappa with more favorable properties with regard to these agreement tables.

In this paper we present several properties of kappa coefficients with symmetric weighting schemes with respect to this particular class of agreement tables. The paper is organized as follows. In the next section we introduce notation, define weighted kappa, and discuss some of its special cases, including Cohen's kappa. The results are presented in Section 3. Section 4 contains a conclusion.

2. Kappa Coefficients

In this section we introduce notation and define the kappa coefficients. For notational convenience weighted kappa is here defined in terms of dissimilarity scaling [8]. If the weights are dissimilarities, pairs of categories that are further apart are assigned higher weights.

Suppose two fixed observers independently rate the same set of n subjects using the same set of $c \geq 2$ categories that are defined in advance. For a population of subjects, let π_{ij} denote the proportion classified in category i by the first observer and in category j by the second observer, where $1 \leq i, j \leq c$. The quantities

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

are the marginal probabilities. They reflect how often the observers used the categories. The cell probabilities of the square table $\{\pi_{ij}\}$ are not directly observed. Let $\{n_{ij}\}$ denote the contingency table of observed frequencies. Assuming a multinominal sampling model with the total number of subjects n fixed, the maximum likelihood estimate of π_{ij} is given by $\widehat{\pi}_{ij} = n_{ij}/n$ [14, 15]. Since the rows and columns of $\{n_{ij}\}$ have the same labels, the contingency table is usually called an agreement table. Table 1 presents two hypothetical agreement tables with three categories A, B, and C.

Let $w_{ij} \geq 0$ for $1 \leq i, j \leq c$ be nonnegative real numbers with $w_{ii} = 0$. The weighted kappa coefficient can be defined as [8, 12]

$$\kappa_w = 1 - \frac{\sum_{i=1}^{c} \sum_{j=1}^{c} w_{ij} \pi_{ij}}{\sum_{i=1}^{c} \sum_{j=1}^{c} w_{ij} \pi_{i+} \pi_{+j}}. \tag{2}$$

The numerator of the fraction in (2) is the weighted observed disagreement, while the denominator of the fraction is the weighted chance-expected disagreement. The value of (2) is 1 when there is perfect agreement between the two observers, zero when the weighted observed disagreement is equal to the weighted chance-expected disagreement, and negative when the weighted observed disagreement is larger than the weighted chance-expected disagreement.

Under a multinominal sampling model with n fixed, the maximum likelihood estimate of (2) is

$$\widehat{\kappa}_w = 1 - \frac{n \sum_{i=1}^{c} \sum_{j=1}^{c} w_{ij} n_{ij}}{\sum_{i=1}^{c} \sum_{j=1}^{c} w_{ij} n_{i+} n_{+j}}. \tag{3}$$

Estimate (3) is obtained by substituting $\widehat{\pi}_{ij} = n_{ij}/n$ for the cell probabilities π_{ij} in (2). A large sample standard error of (3) can be found in [16].

In this paper we are interested in the following special case of (2). We may require that weighted kappa has a symmetric weighting scheme; that is, $w_{ij} = w_{ji}$ for $1 \leq i, j \leq c$. Since $w_{ii} = 0$ for $1 \leq i \leq c$, this symmetric kappa is given by

$$\kappa_s = 1 - \frac{\sum_{i=1}^{c-1} \sum_{j=i+1}^{c} w_{ij} \left(\pi_{ij} + \pi_{ji}\right)}{\sum_{i=1}^{c-1} \sum_{j=i+1}^{c} w_{ij} \left(\pi_{i+} \pi_{+j} + \pi_{j+} \pi_{+i}\right)}. \tag{4}$$

Special cases of coefficient (4) that are used in practice are Cohen's kappa [5, 7, 12] for nominal categories and linear kappa [10, 17] and quadratic kappa [9, 11, 18] for ordinal categories. Cohen's kappa and quadratic kappa each have been used in thousands of applications [6, 11, 19]. The two coefficients are briefly discussed below.

The identity weights are defined as

$$w_{ij} = 1_{i \neq j} = \begin{cases} 0 & \text{for } i = j, \\ 1 & \text{for } i \neq j. \end{cases} \tag{5}$$

An example of weighting scheme (5) is presented in the left panel of Table 2. If we use weighting scheme (5) in (2), we obtain Cohen's unweighted kappa [5]

$$\kappa = 1 - \frac{1 - \sum_{i=1}^{c} \pi_{ii}}{1 - \sum_{i=1}^{c} \pi_{i+} \pi_{+i}}. \tag{6}$$

Perhaps a more familiar definition of Cohen's kappa is

$$\kappa = \frac{\sum_{i=1}^{c} \pi_{ii} - \sum_{i=1}^{c} \pi_{i+} \pi_{+i}}{1 - \sum_{i=1}^{c} \pi_{i+} \pi_{+i}}. \tag{7}$$

Formulas (6) and (7) are equivalent; definition (6) will be used in Section 3 below. Coefficient (6) has value 1 when

the observers agree completely, value zero when agreement is equal to that expected under independence, and negative value when agreement is less than expected by chance.

The quadratic weights are defined as $w_{ij} = (i-j)^2$ for $1 \le i, j \le c$. An example of the weights is presented in the right panel of Table 2. If we use the quadratic weights in (2), we obtain the quadratic kappa [9, 18]

$$\kappa_q = 1 - \frac{\sum_{i=1}^{c}\sum_{j=1}^{c}(i-j)^2\pi_{ij}}{\sum_{i=1}^{c}\sum_{j=1}^{c}(i-j)^2\pi_{i+}\pi_{+j}}. \tag{8}$$

Coefficient (8) is the most popular version of weighted kappa in the case that the categories of the rating system are ordinal [2, 11, 19]. The quadratic kappa can be interpreted as an intraclass correlation, which is a proportion of variance [9, 18]. However, the quadratic kappa is not always sensitive to differences in exact agreement [11], and high values of the quadratic kappa can be found even when the level of exact agreement is low [19].

3. A Class of Agreement Tables

It is sometimes desirable to combine some of the categories [7]. For example, when two categories are frequently confused, combining the categories may improve the reliability of the rating system. Suppose we combine two categories i and j, and let $d \ge 0$ be a nonnegative real number. In this paper we focus on the class of agreement tables that satisfy the condition

$$\frac{\pi_{ij}+\pi_{ji}}{\pi_{i+}\pi_{+j}+\pi_{j+}\pi_{+i}} = d \quad \text{for } i \ne j,\ 1 \le i, j \le c. \tag{9}$$

Condition (9) holds, for example, if there is perfect agreement between the raters. In this case $d = 0$ and we have $\sum_{i=1}^{c}\pi_{ii} = 1$ and $\pi_{ij} = 0$ for $i \ne j$ and $1 \le i, j \le c$. It turns out that there are many nonperfect agreement tables that also satisfy (9). Examples are the agreement tables in Table 1. For the two tables, the value of d is .397 and .644, respectively. The examples in Table 1 show that agreement tables that satisfy (9) are not necessarily symmetric. Furthermore, since the examples appear to be ordinary agreement tables that can be encountered in practice, it appears that the class of agreement tables satisfying (9) is not trivial.

For Cohen's kappa in (6) Schouten [13] showed that if (9) holds, then the kappa value cannot be increased or decreased by combing categories. In this section we present various additional results for other special cases of symmetric kappa in (4). Theorem 1 shows that all special cases of symmetric kappa coincide if (9) holds.

Theorem 1. *If* (9) *holds, then* $\kappa_s = 1 - d$.

Proof. If (9) holds, we have the particular case

$$1 - \frac{\pi_{12}+\pi_{21}}{\pi_{1+}\pi_{+2}+\pi_{2+}\pi_{+1}} = 1 - d. \tag{10}$$

Furthermore, for two arbitrary categories i and j with $i \ne j$ we have

$$\begin{aligned}\frac{\pi_{ij}+\pi_{ji}}{\pi_{i+}\pi_{+j}+\pi_{j+}\pi_{+i}} &= \frac{a_{ij}(\pi_{12}+\pi_{21})}{a_{ij}(\pi_{1+}\pi_{+2}+\pi_{2+}\pi_{+1})} \\ &= \frac{\pi_{12}+\pi_{21}}{\pi_{1+}\pi_{+2}+\pi_{2+}\pi_{+1}}\end{aligned} \tag{11}$$

for certain nonnegative real numbers $a_{ij} \ge 0$. Hence, using these a_{ij} and identity (10) we can write κ_s as

$$\begin{aligned}\kappa_s &= 1 - \frac{\sum_{i=1}^{c-1}\sum_{j=i+1}^{c} w_{ij}a_{ij}(\pi_{12}+\pi_{21})}{\sum_{i=1}^{c-1}\sum_{j=i+1}^{c} w_{ij}a_{ij}(\pi_{1+}\pi_{+2}+\pi_{2+}\pi_{+1})} \\ &= 1 - \frac{(\pi_{12}+\pi_{21})\left(\sum_{i=1}^{c-1}\sum_{j=i+1}^{c} w_{ij}a_{ij}\right)}{(\pi_{1+}\pi_{+2}+\pi_{2+}\pi_{+1})\left(\sum_{i=1}^{c-1}\sum_{j=i+1}^{c} w_{ij}a_{ij}\right)} \\ &= 1 - \frac{\pi_{12}+\pi_{21}}{\pi_{1+}\pi_{+2}+\pi_{2+}\pi_{+1}} = 1 - d.\end{aligned} \tag{12}$$

□

A converse version of Theorem 1 also holds. Lemma 2 is used in the proof of Theorem 3.

Lemma 2. *Let* $a, b \ge 0$ *and* $c, d > 0$ *be real numbers. One has*

$$\frac{a}{c} = \frac{b}{d} \Longleftrightarrow \frac{a}{c} = \frac{a+b}{c+d}. \tag{13}$$

Proof. Since c and d are positive numbers, we have $a/c = b/d$ or $ad = bc$. Adding ac to both sides we obtain $a(c+d) = c(a+b)$ or $a/c = (a+b)/(c+d)$. □

Theorem 3. *If all special cases of symmetric kappa are equal, then* (9) *holds.*

Proof. Let $r, r' \in \{1, 2, \ldots, c\}$ with $r \ne r'$ be arbitrary categories. Let κ_s^* denote the value of the special case of symmetric kappa with $w_{rr'} = w_{r'r} = 2$ and all other off-diagonal weights equal to 1. Since all special cases of symmetric kappa are equal, we have in particular $\kappa = \kappa_s^* = 1 - d$ for some real number $d \ge 0$. Using (6), the identity $\kappa = \kappa_s^*$ is equivalent to

$$\begin{aligned}&\frac{1-\sum_{i=1}^{c}\pi_{ii}}{1-\sum_{i=1}^{c}\pi_{i+}\pi_{+i}} \\ &\quad = \frac{1-\sum_{i=1}^{c}\pi_{ii}+\pi_{rr'}+\pi_{r'r}}{1-\sum_{i=1}^{c}\pi_{i+}\pi_{+i}+\pi_{r+}\pi_{+r'}+\pi_{r'+}\pi_{+r}}.\end{aligned} \tag{14}$$

Since $1 - \sum_{i=1}^{c}\pi_{i+}\pi_{+i} > 0$, it follows from application of Lemma 2 to identity (14) and the use of identity (6) that

$$\frac{\pi_{rr'}+\pi_{r'r}}{\pi_{r+}\pi_{+r'}+\pi_{r'+}\pi_{+r}} = \frac{1-\sum_{i=1}^{c}\pi_{ii}}{1-\sum_{i=1}^{c}\pi_{i+}\pi_{+i}} = 1 - \kappa = d. \tag{15}$$

□

Note that in the proof of Theorem 3 certain special cases of coefficient (4) are used. Condition (9) will not necessarily hold if two arbitrary special cases of symmetric kappa are equal. We have the following consequences of Theorems 1 and 3.

Corollary 4. *It holds that* $\kappa_s = 1 \Leftrightarrow \pi_{ij} = 0$ *for* $i \neq j$ *and* $1 \leq i, j \leq c$.

Corollary 5. *It holds that*

$$\kappa_s = 0 \Longleftrightarrow \frac{\pi_{ij} + \pi_{ji}}{\pi_{i+}\pi_{+j} + \pi_{j+}\pi_{+i}} = 1 \quad \text{for } i \neq j, \quad 1 \leq i, j \leq c. \tag{16}$$

Theorem 6 shows that if (9) holds, then the value of coefficient (4) remains constant when we combine two categories.

Theorem 6. *Let* κ_s *denote the value of symmetric kappa of an agreement table with* $c \geq 3$ *categories and* κ_s^* *the value of the table that is obtained by combining categories* r' *and* r''. *If condition* (9) *holds, then one has* $\kappa_s = \kappa_s^*$.

Proof. Since (9) holds, it follows from Theorem 1 that $\kappa_s = 1 - d$ for some $d \geq 0$. Let r denote the category that is obtained by merging r' and r''. Let i with $1 \leq i \leq c$ and $i \neq r', r''$ be an arbitrary category. We have the four relations

$$\pi_{ir} = \pi_{ir'} + \pi_{ir''}, \tag{17a}$$

$$\pi_{ri} = \pi_{r'i} + \pi_{r''i}, \tag{17b}$$

$$\pi_{r+} = \pi_{r'+} + \pi_{r''+}, \tag{17c}$$

$$\pi_{+r} = \pi_{+r'} + \pi_{+r''}. \tag{17d}$$

Furthermore, since (9) holds, we have the identities

$$\frac{\pi_{ir'} + \pi_{r'i}}{\pi_{i+}\pi_{+r'} + \pi_{r'+}\pi_{+i}} = d, \tag{18a}$$

$$\frac{\pi_{ir''} + \pi_{r''i}}{\pi_{i+}\pi_{+r''} + \pi_{r''+}\pi_{+i}} = d. \tag{18b}$$

Applying Lemma 2 to the identities in (18a) and (18b) we obtain

$$\frac{\pi_{ir'} + \pi_{r'i} + \pi_{ir''} + \pi_{r''i}}{\pi_{i+}\pi_{+r'} + \pi_{r'+}\pi_{+i} + \pi_{i+}\pi_{+r''} + \pi_{r''+}\pi_{+i}} = d. \tag{19}$$

Moreover, using (17a), (17b), (17c), (17d), and (19), we have

$$\begin{aligned} &\frac{\pi_{ir} + \pi_{ri}}{\pi_{i+}\pi_{+r} + \pi_{r+}\pi_{+i}} \\ &= \frac{\pi_{ir'} + \pi_{ir''} + \pi_{r'i} + \pi_{r''i}}{\pi_{i+}\left(\pi_{+r'} + \pi_{+r''}\right) + \left(\pi_{r'+} + \pi_{r''+}\right)\pi_{+i}} \\ &= \frac{\pi_{ir'} + \pi_{r'i} + \pi_{ir''} + \pi_{r''i}}{\pi_{i+}\pi_{+r'} + \pi_{r'+}\pi_{+i} + \pi_{i+}\pi_{+r''} + \pi_{r''+}\pi_{+i}} = d. \end{aligned} \tag{20}$$

It follows from identity (20) that condition (9) also holds for the collapsed $(c-1) \times (c-1)$ table. Application of Theorem 1 then yields that $\kappa_s^* = 1 - d$, from which we may conclude that $\kappa_s = \kappa_s^*$. □

Theorem 6 shows that if the value of Cohen's kappa in (6) remains constant when categories are combined, then the value of symmetric kappa in (4) also remains constant when categories are combined. By repeatedly applying Theorem 6 we obtain the following consequence.

Corollary 7. *Let* κ_s *denote the value of symmetric kappa of an agreement table with* $c \geq 3$ *categories and* κ_s^* *the value of the collapsed table corresponding to any partitioning of the categories. If* (9) *holds, then one has* $\kappa_s = \kappa_s^*$.

4. Conclusion

Kappa coefficients are standard tools for summarizing agreement between two observers on a categorical rating scale. The coefficients are nowadays used for summarizing the information in all types of cross-classifications of two variables with the same categories. In the case of nominal categories Cohen's kappa is a standard tool. In this paper we considered a class of agreement tables for which the value of Cohen's kappa remains constant when two categories are combined. It was shown that for this class of agreement tables all special cases of symmetric kappa, that is, all kappa coefficients with a symmetric weighting scheme, coincide (Theorem 1). Furthermore, for this class of agreement tables the value of symmetric kappa remains constant when categories are merged (Theorem 6 and Corollary 7).

Conflict of Interests

The author declares that there is no conflict of interests regarding the publication of this paper.

Acknowledgments

The author thanks an anonymous reviewer for several helpful comments and valuable suggestions on a previous version of the paper. The comments have improved the presentation of the paper. This research is part of Veni project 451-11-026 funded by the Netherlands Organisation for Scientific Research.

References

[1] U. Jakobsson and A. Westergren, "Statistical methods for assessing agreement for ordinal data," *Scandinavian Journal of Caring Sciences*, vol. 19, no. 4, pp. 427–431, 2005.

[2] M. Maclure and W. C. Willett, "Misinterpretation and misuse of the Kappa statistic," *The American Journal of Epidemiology*, vol. 126, no. 2, pp. 161–169, 1987.

[3] A. Agresti, "Modelling patterns of agreement and disagreement," *Statistical Methods in Medical Research*, vol. 1, no. 2, pp. 201–218, 1992.

[4] A. Agresti, *Categorical Data Analysis*, Wiley, Hoboken, NJ, USA, 2002.

[5] J. Cohen, "A coefficient of agreement for nominal scales," *Educational and Psychological Measurement*, vol. 20, pp. 37–46, 1960.

[6] L. M. Hsu and R. Field, "Interrater agreement measures: comments on Kappa_n, Cohen's Kappa, Scott's π, and Aickin's α," *Understanding Statistics*, vol. 2, no. 3, pp. 205–219, 2003.

[7] M. J. Warrens, "Cohen's kappa can always be increased and decreased by combining categories," *Statistical Methodology*, vol. 7, no. 6, pp. 673–677, 2010.

[8] J. Cohen, "Weighted kappa: nominal scale agreement provision for scaled disagreement or partial credit," *Psychological Bulletin*, vol. 70, no. 4, pp. 213–220, 1968.

[9] J. L. Fleiss and J. Cohen, "The equivalence of weighted kappa and the intraclass correlation coefficient as measures of reliability," *Educational and Psychological Measurement*, vol. 33, pp. 613–619, 1973.

[10] S. Vanbelle and A. Albert, "A note on the linearly weighted kappa coefficient for ordinal scales," *Statistical Methodology*, vol. 6, no. 2, pp. 157–163, 2009.

[11] M. J. Warrens, "Some paradoxical results for the quadratically weighted kappa," *Psychometrika*, vol. 77, no. 2, pp. 315–323, 2012.

[12] M. J. Warrens, "Conditional inequalities between Cohen's kappa and weighted kappas," *Statistical Methodology*, vol. 10, pp. 14–22, 2013.

[13] H. J. A. Schouten, "Nominal scale agreement among observers," *Psychometrika*, vol. 51, no. 3, pp. 453–466, 1986.

[14] A. Agresti, *Categorical Data Analysis*, Wiley, New York, NY, USA, 1990.

[15] Y. M. M. Bishop, S. E. Fienberg, and P. W. Holland, *Discrete Multivariate Analysis: Theory and Practice*, The MIT Press, Cambridge, Mass, USA, 1975.

[16] J. L. Fleiss, J. Cohen, and B. S. Everitt, "Large sample standard errors of kappa and weighted kappa," *Psychological Bulletin*, vol. 72, no. 5, pp. 323–327, 1969.

[17] D. Cicchetti and T. Allison, "A new procedure for assessing reliability of scoring EEG sleep recordings," *The American Journal of EEG Technology*, vol. 11, pp. 101–109, 1971.

[18] 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*, vol. 64, no. 2, pp. 243–253, 2004.

[19] P. Graham and R. Jackson, "The analysis of ordinal agreement data: beyond weighted kappa," *Journal of Clinical Epidemiology*, vol. 46, no. 9, pp. 1055–1062, 1993.