

University of Groningen

Symmetric kappa as a function of unweighted kappas

Warrens, Matthijs

Published in:
Communications in Statistics. Part B: Simulation and Computation

DOI:
[10.1080/03610918.2016.1146765](https://doi.org/10.1080/03610918.2016.1146765)

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version
Publisher's PDF, also known as Version of record

Publication date:
2017

[Link to publication in University of Groningen/UMCG research database](#)

Citation for published version (APA):
Warrens, M. J. (2017). Symmetric kappa as a function of unweighted kappas. Communications in Statistics. Part B: Simulation and Computation, 46(7), 5240-5245. DOI: 10.1080/03610918.2016.1146765

Copyright

Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

Take-down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): <http://www.rug.nl/research/portal>. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.



Symmetric kappa as a function of unweighted kappas

Matthijs J. Warrens

GION, University of Groningen, Groningen, The Netherlands

ABSTRACT

It is shown that a symmetric kappa corresponding to a $c \times c$ table with $c \geq 3$ categories can be written as a function of the unweighted kappa corresponding to the same table and the $c(c - 1)/2$ distinct unweighted kappas associated with the $(c - 1) \times (c - 1)$ tables that are obtained by combining two categories. The result is a new MGB-type result.

ARTICLE HISTORY

Received 14 November 2015
Accepted 18 January 2016

KEYWORDS

Combining categories;
Interobserver agreement;
Symmetric kappa;
Unweighted kappa;
Weighted kappa

MATHEMATICS SUBJECT CLASSIFICATION

62H20; 62H17; 62P25

1. Introduction

In many applications in the social, behavioral, and biomedical science, rating systems and scales are used to classify subjects and objects into mutually exclusive categories. Examples are, a pathologist that rates the severity of lesions from scans, a psychologist that assesses the degree of anxiety of a client, and competing diagnostic devices that classify the extent of a disease in patients. Categorical ratings usually entail a certain degree of subjective judgment. Therefore, an important issue is assessment of the reliability of the categorical rating system. One way to assess reliability is to let two raters classify independently the same set of subjects with the rating system. The reliability of the system can then be assessed by analyzing the agreement between the observers. If there is high agreement between the ratings there is consensus in the diagnosis and the ratings can be considered interchangeable.

Researchers usually express agreement between raters in a single number. Commonly used coefficients for doing this are Cohen's unweighted kappa in the case of nominal (unordered) categories (Warrens, 2013a, 2015a; Yang and Zhou, 2014) and weighted kappa in the case of ordered categories (Warrens, 2012a, 2014; Yang and Zhou, 2015). Both kappa coefficients correct for agreement due to chance. Not all disagreements may be equally important. Weighted kappa allows the use of weights to take the importance of the disagreements between the categories into account. For example, with ordered categories one usually expects more disagreement or confusion between adjacent categories than on categories that are further apart.

A criticism frequently formulated against the use of weighted kappa is that the weights are arbitrarily defined (Vanbelle and Albert, 2009; Warrens, 2012a). To find support for certain weighting schemes several authors have studied algebraic and analytic properties of variants of

CONTACT Matthijs J. Warrens m.j.warrens@rug.nl GION, University of Groningen, Grote Rozenstraat 3, 9712 TG Groningen, The Netherlands.

Published with license by Taylor & Francis. © 2017 Matthijs J. Warrens

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

weighted kappa. For example, kappa with quadratic weights can be interpreted as an intraclass correlation coefficient (Schuster, 2004). Furthermore, squared weights can be chosen such that weighted kappa is equivalent to another intraclass correlation or the Pearson correlation (Warrens, 2014). The latter result suggests that for rating systems with ordinal categories we may abandon the weighted kappa altogether and replace it with the Pearson correlation.

Several authors have also found support for linear kappa, i.e., weighted kappa with linear weights. In each case, it was shown that the kappa corresponding to the original contingency table is a function of kappas corresponding to smaller tables that can be obtained by merging adjacent categories of the original table. Vanbelle and Albert (2009) showed that the components of linear kappa can be obtained from all distinct 2×2 tables. In fact, linear kappa is a weighted average of the 2×2 kappas, where the weights are the denominators of the 2×2 kappas (Warrens, 2011a). This result was generalized by Warrens (2012b) who showed that the overall linear kappa is a weighted average of the linear kappas corresponding to all distinct smaller tables of a specific size that can be obtained by merging adjacent categories. Warrens (2013b, 2015b) showed that the properties of linear kappa extend to a more general kappa coefficient called additive kappa, i.e., weighted kappa with additive weights. With additive weights, the categories have an underlying one-dimensional interval scale but are not required to be equally spaced.

The properties of linear and additive kappa preserves, in a way, an analogous property of unweighted kappa. With ordered categories it only makes sense to combine adjacent categories. With unordered categories, we may combine all the categories in different ways. Warrens (2011b) showed that given a partition type of the categories, the overall kappa is a weighted average of the kappas of the collapsed tables corresponding to all partitions of that type. The weights are the denominators of the kappas corresponding to the collapsed tables.

In the above-mentioned research on unweighted, linear, and additive kappa it was shown that an overall kappa coefficient is a function of the same type of kappa coefficients associated with smaller tables. A completely new type of result was presented in Moradzadeh et al. (2017), or MGB for short. These authors showed that certain weighted kappas are functions of unweighted kappas. The unweighted kappas correspond to smaller tables that can be obtained by combining adjacent categories. A requirement with regard to the weighting schemes is that pairs of categories that are the same number of steps apart have identical weights. Examples are the linear and quadratic kappa.

In this article, we present a new MGB-type result. It is shown that a symmetric kappa corresponding to a $c \times c$ table with $c \geq 3$ categories can be written as a function of the unweighted kappa corresponding to the same table and the $c(c - 1)/2$ distinct unweighted kappas associated with the $(c - 1) \times (c - 1)$ tables that are obtained by combining two categories. The article is organized as follows. Definitions are presented in Section 2 and the main result is presented in Section 3. In Section 4, it is shown, in the context of a rating system with three categories, how the above-mentioned results and the new result are related. Section 5 contains a conclusion.

2. Weighted and unweighted kappa

In this section, we define Cohen's weighted kappa and unweighted kappa. Suppose two raters classify independently the same group of N subjects into one of $c \geq 2$ categories. The categories are defined in advance and the raters use the same c categories. Let π_{ij} for $i, j = 1, 2, \dots, c$ denote the proportion of subjects classified into category i by the first rater and category j by the second rater. The square contingency table $\{\pi_{ij}\}$ reflects the joint agreement

between the two raters. The marginal totals are denoted by π_{i+} for the first rater and π_{+i} for the second rater. The totals reflect how often the raters have used the categories.

Let the weight w_{ij} be a real number with $0 \leq w_{ij} \leq 1$ for $i, j = 1, 2, \dots, c$. We set $w_{ij} = 1$ if $i = j$. Hence, the elements on the main diagonal of the contingency table $\{\pi_{ij}\}$ get the maximum weight of 1. Commonly used weights are the linear weights (Vanbelle and Albert, 2009; Warrens, 2011a, 2012b) and quadratic weights (Schuster, 2004; Warrens, 2012a, 2013a). Weighted kappa can be defined as

$$\kappa_w = \frac{O_w - E_w}{1 - E_w} = \frac{\sum_{i=1}^c \sum_{j=1}^c w_{ij}(\pi_{ij} - \pi_{i+}\pi_{+j})}{1 - \sum_{i=1}^c \sum_{j=1}^c w_{ij}\pi_{i+}\pi_{+j}}, \quad (1)$$

where

$$O_w = \sum_{i=1}^c \sum_{j=1}^c w_{ij}\pi_{ij} \quad (2)$$

is the weighted observed agreement and

$$E_w = \sum_{i=1}^c \sum_{j=1}^c w_{ij}\pi_{i+}\pi_{+j} \quad (3)$$

is the weighted expected agreement. To avoid indeterminate case, we assume that $E_w < 1$. The kappa coefficient in (1) has value 1 if the weighted observed agreement equals unity ($O_w = 1$) and value 0 if the weighted observed agreement is equal to the weighted expected agreement ($O_w = E_w$).

If we use the identity weights

$$w_{ij} = \begin{cases} 1, & \text{if } i = j, \\ 0, & \text{otherwise,} \end{cases} \quad (4)$$

in (1), we obtain Cohen's unweighted kappa given by

$$\kappa = \frac{O - E}{1 - E} = \frac{\sum_{i=1}^c (\pi_{ii} - \pi_{i+}\pi_{+i})}{1 - \sum_{i=1}^c \pi_{i+}\pi_{+i}}, \quad (5)$$

where

$$O = \sum_{i=1}^c \pi_{ii} \quad (6)$$

is the raw observed agreement and

$$E = \sum_{i=1}^c \pi_{i+}\pi_{+i} \quad (7)$$

is the expected agreement. The kappa coefficient in (5) has value 1 if there is perfect observed agreement between the raters ($O = 1$) and value 0 if the observed agreement is equal to the agreement expected under independence ($O = E$).

3. A theorem

In this section, it is shown that weighted kappa with symmetric weights can be expressed as a function of unweighted kappas. In the case of symmetric kappa the (i, j) th cell and (j, i) th

cell of the contingency table $\{\pi_{ij}\}$ have the same weight, i.e., $w_{ij} = w_{ji}$. The commonly used unweighted kappa, linear kappa, and quadratic kappa are examples of symmetric kappa.

Since $w_{ii} = 1$ for all i and $w_{ij} = w_{ji}$ for all i and j , symmetric kappa can be written as

$$\kappa_s = \frac{O - E + \sum_{i=1}^{c-1} \sum_{j=i+1}^c w_{ij}(\pi_{ij} + \pi_{ji} - (\pi_{i+}\pi_{+j} + \pi_{j+}\pi_{+i}))}{1 - E - \sum_{i=1}^{c-1} \sum_{j=i+1}^c w_{ij}(\pi_{i+}\pi_{+j} + \pi_{j+}\pi_{+i})}. \tag{8}$$

In **Theorem 3.1**, we are interested in the observed agreement, expected agreement and unweighted kappa corresponding to the $(c - 1) \times (c - 1)$ that is obtained if categories i and j are combined. The categories can be combined by merging the i th and j th row and the i th and j th column of the square contingency table $\{\pi_{ij}\}$. **Theorem 3.1** shows that for a given $c \times c$ table the corresponding symmetric kappa can be written as a function of the unweighted kappa associated with the $c \times c$ table and the $c(c - 1)/2$ distinct unweighted kappas associated with the $(c - 1) \times (c - 1)$ tables that are obtained by combining two categories.

Theorem 3.1. *Suppose $c \geq 3$. Let O_{ij} and E_{ij} denote, respectively, the observed and expected agreement of the $(c - 1) \times (c - 1)$ that is obtained by combining categories i and j . The associated unweighted kappa is given by*

$$\kappa_{ij} = \frac{O_{ij} - E_{ij}}{1 - E_{ij}}. \tag{9}$$

It holds that

$$\kappa_s = \frac{(1 - \sum_{i=1}^{c-1} \sum_{j=i+1}^c w_{ij})(1 - E)\kappa + \sum_{i=1}^{c-1} \sum_{j=i+1}^c w_{ij}(1 - E_{ij})\kappa_{ij}}{(1 - \sum_{i=1}^{c-1} \sum_{j=i+1}^c w_{ij})(1 - E) + \sum_{i=1}^{c-1} \sum_{j=i+1}^c w_{ij}(1 - E_{ij})}. \tag{10}$$

Proof. Using the identities

$$(1 - E)\kappa = O - E,$$

$$(1 - E_{ij})\kappa_{ij} = O_{ij} - E_{ij},$$

and combining different terms, kappa (10) can be written as

$$\kappa_s = \frac{O - E + \sum_{i=1}^{c-1} \sum_{j=i+1}^c w_{ij}(O_{ij} - O - (E_{ij} - E))}{1 - E - \sum_{i=1}^{c-1} \sum_{j=i+1}^c w_{ij}(E_{ij} - E)}. \tag{11}$$

Next, if we combine categories i and j the observed and expected agreement of the collapsed $(c - 1) \times (c - 1)$ are given by

$$O_{ij} = O + \pi_{ij} + \pi_{ji}, \tag{12}$$

$$E_{ij} = E + \pi_{i+}\pi_{+j} + \pi_{j+}\pi_{+i}. \tag{13}$$

Using these identities in kappa (11), we obtain kappa (8), which completes the proof. \square

Although formula (11) is a simpler expression of symmetric kappa, formula (10) specifies how symmetric kappa is a function of the unweighted kappas. Furthermore, if

$$\sum_{i=1}^{c-1} \sum_{j=i+1}^c w_{ij} \leq 1, \tag{14}$$

all quantities in the numerator and denominator of (10), except the unweighted kappa values, are non-negative. Thus, if inequality (14) holds symmetric kappa is not just a function but a weighted average of the unweighted kappas as well.

4. The 3 × 3 case

In this section, we show in detail how [Theorem 3.1](#) from [Section 3](#) is related to other results from the literature. For notational convenience, we assume that the rating system has $c = 3$ categories, i.e., the contingency table $\{\pi_{ij}\}$ has size 3×3 .

Let κ denote the corresponding unweighted kappa and let κ_{12} , κ_{13} and κ_{23} denote the three unweighted kappas corresponding to the 2×2 tables that are obtained by combining categories 1 and 2, 1 and 3, and 2 and 3, respectively. Furthermore, define

$$w = w_{12} + w_{13} + w_{23}. \quad (15)$$

[Theorem 3.1](#) shows that symmetric kappa can be written as

$$\kappa_s = \frac{(1-w)(1-E)\kappa + w_{12}(1-E_{12})\kappa_{12} + w_{13}(1-E_{13})\kappa_{13} + w_{23}(1-E_{23})\kappa_{23}}{(1-w)(1-E) + w_{12}(1-E_{12}) + w_{13}(1-E_{13}) + w_{23}(1-E_{23})}. \quad (16)$$

Since we have $E, E_{12}, E_{13}, E_{23} < 1$, all weights in (16) are non-negative if $w \leq 1$. In other words, formula (16) shows that symmetric kappa is not only a function, but also a weighted average of the four unweighted kappas if and only if $w < 1$.

Moradzadeh et al. (2017) considered the case where $w_{12} = v = w_{23}$ and $w_{13} = 0$. In this case, symmetric kappa is given by

$$\kappa_s = \frac{(1-2v)(1-E)\kappa + v(1-E_{12})\kappa_{12} + v(1-E_{23})\kappa_{23}}{(1-2v)(1-E) + v(2-E_{12}-E_{23})}. \quad (17)$$

All weights in (17) are non-negative if $v \leq \frac{1}{2}$. In other words, formula (17) shows that symmetric kappa is not only a function, but also a weighted average of the three unweighted kappas if and only if $v < \frac{1}{2}$.

Formulas (16) and (17) shows that the result in Moradzadeh et al. (2017) is a special case of [Theorem 3.1](#) in [Section 3](#) if the rating system has precisely three categories. If the system has four or more categories the result in Moradzadeh et al. (2017) and [Theorem 3.1](#) are different.

Next, if the weights w_{12} , w_{13} and w_{23} in (16) are chosen such that $w = 1$ symmetric kappa becomes

$$\kappa_s = \frac{w_{12}(1-E_{12})\kappa_{12} + w_{13}(1-E_{13})\kappa_{13} + w_{23}(1-E_{23})\kappa_{23}}{w_{12}(1-E_{12}) + w_{13}(1-E_{13}) + w_{23}(1-E_{23})}. \quad (18)$$

Formula (18) shows that symmetric kappa is a weighted average of the three 2×2 kappas only if $w = 1$. If we also require that $w_{13} = 0$ symmetric kappa becomes additive kappa (Warrens, 2013b, 2015b). The latter kappa is given by

$$\kappa_a = \frac{w_{12}(1-E_{12})\kappa_{12} + w_{23}(1-E_{23})\kappa_{23}}{w_{12}(1-E_{12}) + w_{23}(1-E_{23})}. \quad (19)$$

Formula (19) shows that additive kappa is the unique kappa that is a weighted average of kappas κ_{12} and κ_{23} .

Finally, we may require that the distance between categories 1 and 2 and between categories 2 and 3 is equal, i.e., $w_{12} = w_{23}$. If we use this condition in (19) additive kappa reduces to linear kappa (Warrens, 2011a, 2012b). The latter kappa is given by

$$\kappa_\ell = \frac{(1-E_{12})\kappa_{12} + (1-E_{23})\kappa_{23}}{2-E_{12}-E_{23}}. \quad (20)$$

5. Conclusion

In this article, we presented a new MGB-type result. The result holds for any weighted kappa that has a symmetric weighting scheme. A weighting scheme is symmetric if the (i, j) th cell and (j, i) th cell have the same weight. The result specifies that symmetric kappa corresponding to a $c \times c$ table with $c \geq 3$ categories can be written as a function of the unweighted kappa corresponding to the same table and the $c(c - 1)/2$ distinct unweighted kappas associated with the $(c - 1) \times (c - 1)$ tables that are obtained by combining two categories.

Moradzadeh et al. (2017) also showed that weighted kappa can be written as a function of various unweighted kappas. Their result requires that pairs of categories that are the same number of steps apart have identical weights. This condition is less general than the requirement of symmetry in this article.

ORCID

Matthijs J. Warrens  <http://orcid.org/0000-0002-7302-640X>

References

- Moradzadeh, N., Ganjali, M., Baghfalaki, T. (2017). Weighted kappa as a function of unweighted kappas. *Communications in Statistics. Simulation and Computation*, in press.
- Schuster, C. (2004). A note on the interpretation of weighted kappa and its relations to other rater agreement statistics for metric scales. *Educational and Psychological Measurement* 64:243–253.
- Vanbelle, S., Albert, A. (2009). A note on the linearly weighted kappa coefficient for ordinal scales. *Statistical Methodology* 6:157–163.
- Warrens, M. J. (2011a). Cohen's linearly weighted kappa is a weighted average of 2×2 kappas. *Psychometrika* 76:471–486.
- Warrens, M. J. (2011b). Cohen's kappa is a weighted average. *Statistical Methodology* 8:473–484.
- Warrens, M. J. (2012a). Some paradoxical results for the quadratically weighted kappa. *Psychometrika* 77:315–323.
- Warrens, M. J. (2012b). Cohen's linearly weighted kappa is a weighted average. *Advances in Data Analysis and Classification* 6:67–79.
- Warrens, M. J. (2013a). Conditional inequalities between Cohen's kappa and weighted kappas. *Statistical Methodology* 10:14–22.
- Warrens, M. J. (2013b). 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. (2015a). Five ways to look at Cohen's kappa. *Journal of Psychology & Psychotherapy* 5:197.
- Warrens, M. J. (2015b). Additive kappa can be increased by combining adjacent categories. *International Mathematical Forum* 10:323–328.
- Yang, Z., Zhou, M. (2014). Kappa statistic for clustered matched-pair data. *Statistics in Medicine* 33:2612–2633.
- Yang, Z., Zhou, M. (2015). Weighted kappa statistic for clustered matched-pair ordinal data. *Computational Statistics & Data Analysis* 82:1–18.