doi: 10.14355/jmmf.2013.0204.01

On Association Coefficients, Correction for Chance, and Correction for Maximum Value

Matthijs J. Warrens

Institute of Psychology, Leiden University
Wassenaarseweg 52, 2333 AK Leiden, Netherlands

warrens@fsw.leidenuniv.nl

Abstract

This paper studied correction for chance and correction for maximum value as functions on a space of association coefficients. Various properties of both functions are presented. It is shown that the two functions commute under composition; and that the composed function maps a coefficient and all its linear transformations given the marginal totals to the same coefficient. The results presented in the paper have generalized various results from the literature.

Keywords

Association Coefficients; Chance-corrected Coefficients; Marginal Distributions; Coefficient Space

Introduction

Association coefficients are important tools in data analysis and classification that are used to quantify the degree of association between variables (Zegers 1986a, Albatineh et al. 2006). Individual coefficients can be used for summarizing parts of a research study, while matrices of association coefficients can be used as input for multivariate data analysis techniques like, component analysis and cluster analysis (Gower 1966, Gower and Legendre 1986). Well-known examples of association coefficients are Pearson's product-moment correlation to measure the linear dependence between two continuous variables, the Hubert-Arabie adjusted Rand index for comparing partitions of two different clustering algorithms (Hubert and Arabie 1985, Steinley 2004, Warrens 2008a), and Cohen's kappa for measuring the degree of inter-rater agreement on a categorical scale (Cohen 1960; Bloch and Kraemer 1989; Warrens 2010).

Association coefficients may satisfy certain requirements. Various requirements for coefficients have been discussed in Popping (1983), Zegers (1986b) and Warrens (2008b). Since the choice of an association coefficient should always be considered in the context of the data-analytic problem at hand (Gower and Legende 1986) the requirements can be used as guidelines to select the most appropriate association coefficients. However, it may happen that a coefficient does not satisfy a certain requirement. For these coefficients, corrections have been proposed in the literature. A correction transforms one coefficient into a new coefficient, which then satisfies the desideratum associated with the correction. In this paper, two such corrections are studied as functions on a space of association coefficients, namely correction for chance, and correction for maximum value.

The paper is organized as follows. In Section 2 a coefficient space is defined. This space will be the domain of the correction for chance function and the correction for maximum value function. The coefficients in this paper are association coefficients that summarize the information in a contingency table. Although the correction for chance function has also been used with association coefficients from other data-analytic contexts, correction for maximum value has been studied primarily with coefficients for contingency tables. In Section 3 the correction for chance function is defined and some of its properties are presented. In Section 4 the correction for maximum value function and some of its properties are defined. In Section 5 it is shown that the two functions commute; and that the composed function maps a coefficient and all its linear transformations given the marginal totals to a unique fixed point of the function. In Section 6 it is shown that the correction for chance function and the correction for maximum value function, together with the identity function and their composition, form a commutative idempotent monoid. Section 7 contains a conclusion.

The correction for chance function has been studied by other authors (Albatineh et al. 2006, Warrens 2008b, 2008c, 2011, 2013). All these studies were limited to coefficients that belong to a specific family of linear transformations. In this paper, we presented several new results, and also showed that the results in Albatineh et al. (2006) and Warrens (2008b, 2008c, 2013) hold under more general circumstances.

Association Coefficients

In this section, the coefficient space for the correction for chance function and the correction for maximum value function are defined. In addition, a variety of examples of association coefficients from the literature are presented.

A Coefficient Space

Let $\{p_{ij}\}$ be a contingency table or matrix of size $k \times \ell$ where $k, \ell \geq 2$. It is assumed that the elements of $\{p_{ij}\}$ are non-negative, that is, $p_{ij} \geq 0$ for all i, j, and that the elements of p_{ij} sum to unity. These requirements ensure that the elements p_{ij} are relative frequencies. The quantities

$$p_{i+} = \sum_{j=1}^{\ell} p_{ij} \text{ and } p_{+j} = \sum_{i=1}^{k} p_{ij}$$

will be called the marginal totals of the table $\{p_{ij}\}$. For fixed k and ℓ, consider the set

$$M = \left\{ \{p_{ij}\}_{k\times\ell} \middle| p_{ij} \geq 0 \text{ for all } i,j; \sum_{i,j} p_{ij} = 1 \right\}. \quad (1)$$

The set M consists of all contingency tables of size $k \times \ell$ with non-negative elements that sum to unity. In the context of contingency tables, an association coefficient is a function that assigns to each contingency table a real number. Thus, a coefficient $A: M \to \mathbb{R}$ is a map from the domain M to the real numbers $\mathbb{R}$. For many association coefficients, the codomain is either the closed interval [0,1] or the interval $[-1,1]$. Let $D = \{A: M \to \mathbb{R}\}$ denote the set of all association coefficients from M to $\mathbb{R}$. The coefficient space D will be the domain of the correction for chance function defined in Section 3 and the correction for maximum value function defined in Section 4. In the following subsections, several examples of M and associated elements of D from the literature are available.

Coefficients for 2×2 Tables

Many experimental and research studies can be summarized by a contingency table of size 2×2 (Gower and Legendre 1986, Baulieu 1989, Warrens 2008b, 2008c). This type of table is usually a cross-classification of two binary variables. An example from epidemiology is a reliability study. In a reliability study two observers each rate the same sample of subjects on the presence or absence of a trait or a disease (Fleiss 1975, Bloch and Kraemer 1989). In this case, M is given by

$$M = \left\{ \begin{pmatrix} p_{11} & p_{12} \\ p_{21} & p_{22} \end{pmatrix} \middle| p_{11} + p_{12} + p_{21} + p_{22} = 1 \right\}.$$

Example 1. A well-known association coefficient for 2×2 tables is the phi coefficient

$$\varphi = \frac{p_{11}p_{22} - p_{12}p_{21}}{\sqrt{p_{1+}p_{2+}p_{+1}p_{+2}}} \quad (2)$$

which is the formula of Pearson's product-moment correlation coefficient for two binary variables. Pearson's correlation is widely used as a measure of linear dependence between two variables.

Example 2. Another example is

$$H = \frac{p_{11}p_{22} - p_{12}p_{21}}{min\{p_{1+}p_{+2}, p_{+1}p_{2+}\}}. \quad (3)$$

This coefficient has been discussed in Johnson (1945), but it is better known as Loevinger's H (Loevinger 1947, 1948). It is an important statistic in Mokken scale analysis, a methodology that may be used to select a subset of binary test items that are sensitive to the same underlying dimension (Sijtsma and Molenaar 2002). Coefficient φ can only be equal to 1 if $p_{1+} = p_{+1} = 1/2$, whereas coefficient H can attain its maximum value of unity regardless of the marginal totals.

Coefficients for k×k Tables

In biomedical and behavioral sciences, it is not uncommon to have a research study in which the variables of interest have three or more nominal (unordered) categories. The categories are usually defined beforehand and an experiment may result in two nominal variables with identical categories. An example from developmental psychology is a study in which two coders classify the solution strategies that children use when solving arithmetic problems. In this case, the contingency table is a cross-classification of size $k \times k$ of the two nominal variables. The set M is given by

$$M = \left\{ \{p_{ij}\}_{k\times k} \middle| p_{ij} \geq 0 \text{ for all } i,j; \sum_{i,j} p_{ij} = 1 \right\}.$$

Example 3. The proportion p_{ii} on the main diagonal of a square contingency table reflects how often the two coders agree on category i, or, how often the two variables have category i in the same position. A straightforward coefficient for summarizing agreement is the so-called overall agreement

$$O = \sum_i p_{ii}. \quad (4)$$

Example 4. The most widely used coefficient in biomedical and behavioral science research for summarizing inter-rater agreement on a nominal scale is Cohen's kappa (Cohen 1960, Hanley 1987, Maclure and Willett 1987, Hsu and Field 2003, Warrens 2008a, 2010). The statistic is given by

$$\kappa = \frac{\sum_i (p_{ii} - p_{i+}p_{+i})}{1 - \sum_i p_{i+}p_{+i}}. \quad (5)$$

Unlike the overall agreement O, Cohen's kappa corrects for agreement that may occur due to chance alone. Kappa has value unity if there is perfect agreement, and zero value under statistical independence.

Example 5. A coefficient commonly used in content analysis research is Scott's pi (Scott 1955, Krippendorff 2004a, 2004b). The statistic is defined as

$$\pi = \frac{\sum_i \left(p_{ii} - \frac{1}{4}(p_{i+} + p_{+i})^2\right)}{1 - \frac{1}{4}\sum_i (p_{i+} + p_{+i})^2}. \quad (6)$$

Like Cohen's kappa, Scott's pi corrects for agreement that may occur due to chance alone. The correction is based on different distributional assumptions (see example 10 below).

Example 6. A fourth example is the coefficient

$$H = \frac{\sum_i (p_{ii} - p_{i+}p_{+i})}{\sum_i (min\{p_{i+}, p_{+i}\} - p_{i+}p_{+i})}. \quad (7)$$

This coefficient has been discussed in Cohen (1960), Brennan and Prediger (1981) and Popping (1983). Coefficient (7) generalizes coefficient (3) to the case of three or more nominal categories. Coefficient κ can only be equal to 1 if $p_{i+} = p_{+i}$ for all i, that is, if the marginal distributions of the two variables are identical. Coefficient H can attain its maximum value of unity regardless of the marginal totals.

Cluster Validation Coefficients

In cluster analysis, one is often interested in comparing two partitions of the same set of objects or data points from different clustering algorithms (Steinley 2004, Albatineh et al. 2006, Albatineh and Niewiadomska-Bugaj 2011). In this case, each variable may have different categories. The cluster validation situation closely matches an experiment where two observers each rate the same group of objects using different nominal categories (Hubert 1977, Janson and Vegelius 1982, Popping 1983).

In cluster analysis, the contingency table $\{p_{ij}\}$ is usually called a matching table, and the cell p_{ij} reflects the number of objects placed in cluster i $(i = 1, \ldots, k)$ by the first clustering method and in cluster j $(j = 1, \ldots, \ell)$ by the second method. Association coefficients that summarize the information in a matching table usually compare the number of object pairs placed in the same cluster to the number of object pairs placed in different clusters. There is a formal relationship between these coefficients and association coefficients for 2×2 tables. Let α be the proportion of object pairs placed in the same cluster according to both clustering methods, β (γ) be the proportion of object pairs placed in the same cluster according to one method but not the other method, and δ be the proportion of object pairs not in the same cluster according to either of the methods. The full expressions of α, β, γ and δ in terms of binomial coefficients can be found in Albatineh et al. (2006), Albatineh and Niewiadomska-Bugaj (2011) and Warrens (2008a). The re-parameterisation is given by $\alpha = p_{11}$, $\beta = p_{12}$, $\gamma = p_{21}$ and $\delta = p_{22}$.

Example 7. For some time, the standard tool in cluster analysis for summarizing a matching table was the so-called Rand index (Rand 1971), which is defined as

$$R = \frac{\alpha + \delta}{\alpha + \beta + \gamma + \delta} = \alpha + \delta. \quad (8)$$

Coefficient R compares the number of object pairs placed in the same cluster and in different clusters according to both clustering methods, to the total number of object pairs. In the context of 2×2 tables, this coefficient is called the overall agreement, also known as the simple matching coefficient (Sokal and Michener 1958). In the context of summarizing agreement between two coders that used different nominal categories, the Rand index is equivalent to the coefficient proposed in Brennan and Light (1974).

Example 8. Morey and Agresti (1985) and Hubert and Arabie (1985) argued that the Rand index should be corrected for agreement between the clustering methods due to chance. Nowadays, a standard tool for comparing two partitions of the same objects or data points by two different clustering methods is the Hubert-Arabie adjusted Rand index (Hubert and Arabie 1985, Steinley 2004). Warrens (2008a) showed that in terms of α, β, γ and δ the adjusted Rand index can be written as

$$AR = \frac{2(\alpha\delta - \beta\gamma)}{(\alpha + \beta)(\alpha + \gamma) + (\beta + \delta)(\gamma + \delta)}. \quad (9)$$

In the context of 2×2 tables coefficient, AR is also known as Cohen's kappa (Warrens 2008a).

Linear Transformations

Albatineh et al. (2006) introduced the idea of studying correction for chance for a whole family of validation coefficients simultaneously, and studied coefficients that are linear in $\sum_{i,j} n_{ij}^2$, where n_{ij} is the number of objects placed in cluster i according to the first clustering method and in cluster j according to the second clustering method. Following Albatineh et al. (2006), Warrens (2008b, 2008c, 2011) studied a family of association coefficients for 2×2 tables that have a form $\lambda + \mu(p_{11} + p_{22})$, where $p_{11} + p_{22}$ is the overall agreement and $\lambda = \lambda(p_{1+}, p_{2+}, p_{+1}, p_{+2})$ and $\mu = \mu = (p_{1+}, p_{2+}, p_{+1}, p_{+2})$ are functions of the marginal totals. Lemma 1 shows that if one coefficient is linear in a second coefficient given the marginal totals, then the second coefficient is also linear in the first coefficient.

Lemma 1. Let $A, B \epsilon D$ and suppose $B = \lambda + \mu A$ where λ and μ are functions of the marginal totals. Then $A = \lambda^* + \mu^* B$ where

$$\lambda^* = -\frac{\lambda}{\mu} \text{ and } \mu^* = \frac{1}{\mu}.$$

Since λ and μ are functions of the marginal totals, and λ^* and μ^* are ratios of λ and μ, it follows that λ^* and μ^* in Lemma 1 are also functions of the marginal totals.

Example 9. The phi coefficient φ (Example 1) can be written as $\lambda + \mu O = \lambda + \mu(p_{11} + p_{22})$ where

$$\lambda = \frac{-p_{1+}p_{+1} - p_{2+}p_{+2}}{2\sqrt{p_{1+}p_{2+}p_{+1}p_{+2}}} \text{ and } \mu = \frac{1}{2\sqrt{p_{1+}p_{2+}p_{+1}p_{+2}}}.$$

Vice versa, the overall agreement O can be written as $\lambda^* + \mu^* \varphi$ where

$$\lambda^* = p_{1+}p_{+1} + p_{2+}p_{+2} \text{ and } \mu^* = 2\sqrt{p_{1+}p_{2+}p_{+1}p_{+2}}.$$

Correction for Chance

In this section, the correction for chance function is defined and studied. In several data-analytic contexts, it is desirable that the theoretical value of an association coefficient is zero if the two variables are statistically independent (Popping 1983, Zegers 1986a). The adjusted Rand index and Cohen's kappa each have zero value under independence, but the proportion of overall agreement does not. If a measure does not have zero value under statistical independence, it may be corrected for association due to chance (Fleiss 1975, Krippendorff 1987, Albatineh et al. 2006, Warrens 2008b). Let $E(A)$ denote the value of coefficient A under chance conditionally upon fixed marginal totals, and $M(A)$ denote the overall maximum value of coefficient A. It is assumed that the chance process is such that the expectation $E(A)$ is only a function of the marginal totals. Furthermore, for many association coefficients from the literature we have $M(A) = 1$. The correction for chance function is defined as

$$c: D \to D, A \mapsto \frac{A - E(A)}{M(A) - E(A)},$$

or simplied

$$c(A) = \frac{A - E(A)}{M(A) - E(A)}. \quad (10)$$

The numerator of (10) is the difference between A and $E(A)$, whereas the denominator of (10) is the maximum possible value of the numerator. It is assumed that $M(A)$ is greater than $E(A)$ to avoid indeterminate cases. Different distributional assumptions lead to different definitions of the expectation $E(A)$, and thus different versions of the function c. For cluster validation coefficients, two distributional assumptions have been discussed in Albatineh et al. (2006) and Albatineh and Niewiadomska-Bugaj (2011). Example 10 considers two assumptions for coefficients for $k \times k$ tables.

Example 10. Consider the overall agreement O (Example 3). Suppose that $\{p_{ij}\}$ is a product of chance concerning two different frequency distributions, one for the row categories and the otherfor the column categories (Krippendorff 1987, Warrens 2010). In this case we have

$$E(O) = \sum_i E(p_{ii}) = \sum_i p_{i+}p_{+i}. \quad (11)$$

Expectation (11) is the value of O under statistical independence. Using O, $M(O) = 1$ and $E(O)$ in (11) in (10), we obtain Cohen's kappa (Example 4). Alternatively, it may be assumed that the frequency distribution underlying the row and column categories is the same for the rows and columns (Scott 1955, Krippendorff 1987). We have, for example,

$$E(O) = \sum_i E(p_{ii}) = \sum_i \left(\frac{p_{i+} + p_{+i}}{2}\right)^2. \quad (12)$$

Using O, $M(O) = 1$ and $E(O)$in (12) in (10), we obtain Scott's pi (Example 5).

The function c in (10) has been applied to association coefficients for metric scales (Zegers 1986a, 1986b), coefficients for inter-rater agreement (Zegers 1991, Warrens 2010) and coefficients of cluster validation (Albatineh et al. 2006, Albatineh and Niewiadomska-Bugaj 2011). It has been demonstrated by many authors that association coefficients may become equivalent after correction (10) (Fleiss 1975, Zegers 1986b, Albatineh et al. 2006, Warrens 2008b, 2008c, 2011). These relations show how various association coefficients from the literature are related, and usually provide new ways to interpret the chance-corrected association coefficients.

In the remainder of this section, we study the function

c in the context of the general coefficient space. In the results below, we do not assume a specific form for the expectation $E(A)$. The following lemmas provide some specific conditions for two coefficients to coincide after correction for chance. If a result generalizes an existing result in the literature, the latter result is explicitly mentioned.

Let A be a coefficient, and a and $b \neq 0$ be real numbers. Lemma 2 shows that A and the linear transformation $a + bA$ coincide after correction for chance. The lemma generalizes Proposition 2 in Warrens (2008b).

Lemma 2. Let $A \in D$ and $B = a + bA$, where $a, b \in \mathbb{R}$ are constants with $b \neq 0$. Then $c(A) = c(B)$.

Proof: The definition of $c(A)$ is presented in (10). Since a and b are constants, we have $E(B) = E(a + bA) = a + bE(A)$ and $M(B) = M(a + bA) = a + bM(A)$. Hence, we have

$$c(B) = \frac{a + bA - a - bE(A)}{a + bM(A) - a - bE(A)} = \frac{A - E(A)}{M(A) - E(A)} = c(A).$$

■

Example 11. For a fixed number of categories k, the coefficient given by

$$S = \frac{O - \frac{1}{k}}{1 - \frac{1}{k}} = -\frac{1}{k-1} + \frac{k}{k-1}O, \tag{13}$$

is a linear transformation of the overall agreement O (Warrens 2012). Coefficient S first proposed in Bennett et al. (1954), is equivalent to coefficient C in Janson and Vegelius (1979, p. 260) and coefficient RE proposed in Janes (1979). In Brennan and Prediger (1981) coefficient S is denoted by κ_n. Since $M(O) = 1$ it follows from Lemma 2 that

$$c(O) = c(S) = \frac{O - E(O)}{1 - E(O)}.$$

Lemma 3 considers a condition for the equivalence of a coefficient A and a linear transformation of A.

Lemma 3. Let $A \in D$ and $B = \lambda + \mu A$, where λ and $\mu \neq 0$ are functions of the marginal totals. Then $c(A) = c(B) \Leftrightarrow M(B) = \lambda + \mu M(A)$.

Proof: Since $E(\lambda + \mu A) = \lambda + \mu E(A)$, we have

$$c(A) = c(B) \Leftrightarrow \frac{A - E(A)}{M(A) - E(A)} = \frac{\lambda + \mu A - \lambda - \mu E(A)}{M(B) - \lambda - \mu E(A)}$$
$$\Leftrightarrow \frac{1}{M(A) - E(A)} = \frac{\mu}{M(B) - \lambda - \mu E(A)}$$
$$\Leftrightarrow M(B) = \lambda + \mu M(A).$$

■

A consequence of Lemma 3 is that two linear transformations of a coefficient A coincide if they have the same ratio

$$\frac{M(B) - \lambda}{\mu}. \tag{14}$$

Corollary 4 generalizes a result in Albatineh et al. (2006).

Corollary 4. Let $A \in D$ and $B_1 = \lambda_1 + \mu_1 A$ and $B_2 = \lambda_2 + \mu_2 A$ where λ_1, λ_2, $\mu_1 \neq 0$ and $\mu_2 \neq 0$ are functions of the marginal totals. Then

$$c(B_1) = c(B_2) \Leftrightarrow \frac{M(B_1) - \lambda_1}{\mu_1} = \frac{M(B_2) - \lambda_2}{\mu_2}.$$

Example 12. For the overall agreement O (Example 3) we have $M(O) = 1$. Thus, ratio (14) is given by

$$\frac{M(O) - \lambda_O}{\mu_O} = 1.$$

We can write Cohen's kappa (Example 4) as $\kappa = \lambda_\kappa + \mu_\kappa O$ where

$$\lambda_\kappa = \frac{-\sum_i p_{i+}p_{+i}}{1 - \sum_i p_{i+}p_{+i}} \text{ and } \mu_\kappa = \frac{1}{1 - \sum_i p_{i+}p_{+i}}.$$

Furthermore, since $M(\kappa) = 1$, ratio (14)

$$\frac{M(\kappa) - \lambda_\kappa}{\mu_\kappa} = 1 - \sum_i p_{i+}p_{+i} + \sum_i p_{i+}p_{+i} = 1.$$

We can write Scott's pi (Example 5) as $\pi = \lambda_\pi + \mu_\pi O$ where

$$\lambda_\pi = \frac{-\sum_i \frac{1}{4}(p_{i+} + p_{+i})^2}{1 - \frac{1}{4}\sum_i (p_{i+} + p_{+i})^2} \text{ and } \mu_\pi = \frac{1}{1 - \frac{1}{4}\sum_i (p_{i+} + p_{+i})^2}.$$

Furthermore, since $M(\pi) = 1$, ratio (14)

$$\frac{M(\pi) - \lambda_\pi}{\mu_\pi} = 1.$$

It then follows from Corollary 4, together with Example 11, that O, S, κ and π coincide after correction for chance.

Lemma 5 shows that if two coefficients coincide after correction for chance, then the chance-corrected sum of the coefficients is identical to the individual chance-corrected coefficients. Lemma 5 generalizes Theorem 1 in Warrens (2008b).

Lemma 5. Let $A, B \in D$ with $c(A) = c(B)$. Then $c(A + B) = c(A) = c(B)$.

Proof: Since E and M are linear operators, we have

$$c(A + B) = \frac{A + B - E(A) - E(B)}{M(A) + M(B) - E(A) - E(B)}.$$

Furthermore, using (10), we have the identities

$$A - E(A) = [M(A) - E(A)]c(A)$$
$$B - E(B) = [M(B) - E(B)]c(B).$$

Hence, $c(A + B)$ is equal to

$$\frac{[M(A) - E(A)]c(A) + [M(B) - E(B)]c(B)}{M(A) - E(A) + M(B) - E(B)}. \tag{15}$$

Since $M(A)$ is greater than $E(A)$, the quantity $M(A) - E(A)$ is positive. Equation (15) shows that $c(A+B)$ is a weighted average of $c(A)$ and $c(B)$ with positive weights $M(A) - E(A)$ and $M(B) - E(B)$. Since $c(A) = c(B)$, (15) can be written as

$$c(A+B) = \frac{[M(A) - E(A) + M(B) - E(B)]c(A)}{M(A) - E(A) + M(B) - E(B)} = c(A).$$

■

The correction for chance function (10) defines a relation on D. Two coefficients A and B may be called equivalent with respect to c, denoted by $A \sim B$, if $c(A) = c(B)$. It can be verified that $\sim$ defines an equivalence relation on D. For two coefficients A and B we usually have $c(A+B) \neq c(A) + c(B)$. Thus, in general, c is not a linear map.

Correction for Maximum Value

For many association coefficients, the maximal attainable value is restricted by the marginal totals of the contingency table. For example, the phi coefficient (Example 1) can only be equal to 1 if $p_{1+} = p_{+1} = 1/2$. In the literature, it has been suggested to replace the phi coefficient by the ratio phi/phimax, where phimax is the maximum value of the phi coefficient given the marginal probabilities. A detailed review of the phi/phimax literature is presented in Davenport and El-Sanhurry (1991).

It may be desirable that an association coefficient has maximum value unity regardless of the marginal distributions. An example of a coefficient with this property is coefficient H (Examples 2 and 6). For association coefficients that do not possess this property, the following correction has been suggested (Warrens 2008b). Let $m(A)$ denote the maximum value of coefficient A given the marginal totals. The correction for maximum value function is defined as

$$d: D \to D, A \mapsto \frac{A}{m(A)},$$

or simplied

$$d(A) = \frac{A}{m(A)}. \tag{16}$$

Note that we have the inequality $m(A) \leq M(A)$.

Example 13. Suppose $\{p_{ij}\}$ is a cross-classification of two nominal variables with identical categories. The value of the diagonal element p_{ii} cannot exceed the minimum of p_{i+} and p_{+i}. The maximum value of the overall agreement O (Example 3) is thus restricted by the marginal totals. Its maximum value given the marginal totals is

$$m(O) = \sum_i min\{p_{i+}, p_{i+}\}. \tag{17}$$

Using O and $m(O)$ in the correction for maximum value function (16) we obtain

$$d(O) = \frac{\sum_i p_{ii}}{\sum_i min\{p_{i+}, p_{i+}\}}.$$

Example 14. Using (17) the maximum value of Cohen's kappa given the marginal totals is

$$m(\kappa) = \frac{\sum_i (min\{p_{i+}, p_{i+}\} - p_{i+}p_{+i})}{1 - \sum_i p_{i+}p_{+i}}.$$

Using κ and $m(\kappa)$ in the correction for maximum value function (16), we obtain coefficient H in (7).

In the remainder of this section, the function d in the context of the general coefficient space is studied. The following lemmas provide some specific conditions for two coefficients to coincide after correction for maximum value.

Let A be a coefficient, λ a function of the marginal totals, and let $B = \lambda A$. Lemma 6 shows that A and B coincide after correction for maximum value.

Lemma 6. Let $A \in D$ and $B = \lambda A$, where $\lambda \neq 0$ is a function of the marginal totals. Then $d(A) = d(B)$.

Proof: The formula for $d(A)$ is presented in equation (16). To determine $m(B)$it may be assumed that the marginal totals are given. We have $m(B) = m(\lambda A) = \lambda m(A)$. Hence,

$$d(B) = \frac{\lambda A}{\lambda m(A)} = d(A).$$

■

Lemma 7 shows that if two coefficients coincide after correction for maximum value, then the corrected sum of the coefficients is identical to the individual corrected coefficients.

Lemma 7. Let $A, B \in D$ with $d(A) = d(B)$. Then

$$d(A+B) = d(A) = d(B).$$

Proof: Since m is a linear operator we have

$$d(A+B) = \frac{A+B}{m(A) + m(B)}.$$

Furthermore, using (16), we have the identities $A = m(A)d(A)$ and $B = m(B)d(B)$. Hence,

$$d(A+B) = \frac{m(A)d(A) + m(B)d(B)}{m(A) + m(B)}. \tag{18}$$

The right-hand side of (18) shows that $d(A+B)$ is a weighted average of $d(A)$ and $d(B)$ with positive weights $m(A)$ and $m(B)$. Since $d(A) = d(B)$, (18) can be written as

$$d(A+B) = \frac{[m(A) + m(B)]d(A)}{m(A) + m(B)} = d(A).$$

■

The correction for maximum value function (16) defines a relation on D. Two coefficients A and B may be called equivalent with respect to d if $d(A) = d(B)$. It can be verified that this defines an equivalence relation on D.

Commutative Functions

In this section, the composition of the correction for chance function (10) and the correction for maximum value function (16) are studied. If a coefficient is first corrected for maximum value and then corrected for chance, the composition $c \circ d = cd$ is taken into consideration. If we first correct for chance and then for maximum value, we have the composition $d \circ c = dc$. Theorem 8 shows that the two compositions are equivalent. In other words, the functions c and d commute.

Theorem 8. $cd = dc$.

Proof: Let $A \epsilon D$. $E(A)$ and $m(A)$ are determined. Both quantities require that the marginal totals are given. Hence, let the marginal totals be fixed. We first determine the expression of $(d \circ c)(A) = dc(A)$. The formula for $c(A)$ is given in (10). Since $E(A)$ is a function of the marginal totals, and since the marginal totals are fixed, $E(A)$ is fixed. Furthermore, $M(A)$ is a real constant. Hence,

$$m(c(A)) = \frac{m(A) - E(A)}{M(A) - E(A)}. \tag{19}$$

Dividing (10) by (19) we obtain

$$d\big(c(A)\big) = \frac{c(A)}{m(c(A))} = \frac{A - E(A)}{m(A) - E(A)}. \tag{20}$$

Next, we determine an expression for $(c \circ d)(A) = cd(A)$. Since $m(A)$ is fixed with fixed marginals, we have

$$E(d(A)) = \frac{E(A)}{m(A)}. \tag{21}$$

Furthermore, by definition of (16) we have

$$M\big(d(A)\big) = M\left(\frac{A}{m(A)}\right) = 1. \tag{22}$$

Using (21) and (22) in (10), and multiplying all terms of the result by $m(A)$, we obtain

$$c\big(d(A)\big) = \frac{A - E(A)}{m(A) - E(A)}. \tag{23}$$

Since the right-hand sides of (20) and (23) are identical, we have $d\big(c(A)\big) = c\big(d(A)\big)$.■

Theorem 8 shows that after correction for chance and maximum value coefficient A has a form (20)=(23), regardless of the order in which the corrections are applied. It turns out that formula (20)=(23) has another interesting property. Theorem 9 shows that any linear transformation $\lambda + \mu A$ of a coefficient A, where λ and μ are functions of the marginal totals, coincide with A after correction for both chance and maximum value. In other words, the function cd maps a coefficient A and all its linear transformations to the same coefficient. Theorem 9 generalizes the main result in Warrens (2008c).

Theorem 9. Let $A \epsilon D$ and let $B = \lambda + \mu A$, where λ and $\mu \neq 0$ are functions of the marginal totals. Then $cd(A) = cd(B)$.

Proof: Since both $E(B)$ and $m(B)$ need to be determined, it may be assumed that the marginal totals are fixed. Then $E(B) = \lambda + \mu E(A)$ and $m(B) = \lambda + \mu m(A)$ and it follows that

$$cd(B) = \frac{\lambda + \mu A - \lambda - \mu E(A)}{\lambda + \mu m(A) - \lambda - \mu E(A)} = cd(A).$$

■

Example 15. For the overall agreement O (Example 3) the quantity $m(O)$ is given in (17). It follows from Theorem 9 that any linear transformation $\lambda + \mu O$ of O, where λ and $\mu \neq 0$ are functions of the marginal totals, becomes

$$\frac{O - E(O)}{\sum_i min\{p_{i+}, p_{i+}\} - E(O)}, \tag{24}$$

after correction for chance and maximum value. Using $E(O)$ in (11) in (24) we obtain coefficient H (Example 6).

An Idempotent Commutative Monoid

Theorem 8 from the previous section shows that the functions c and d commute under composition. The identity function on D is given by $1: D \to D, A \mapsto A$. The functions c, d and cd also commute with the identity. In this section, we investigate the algebraic structure of the set $\{1, c, d, cd\}$ under composition. Lemmas 10, 11 and 12 show that the functions c , d and cd are idempotent.

Lemma 10. $c^2 = c$

Proof: Since $M(A)$ is a real number and $E(A)$ a function of the marginal totals, we have

$$E(c(A)) = \frac{E(A) - E(A)}{M(A) - E(A)} = 0.$$

By definition of (10) it is held that $M(c(A)) = 1$. Hence, we have

$$c(c(A)) = \frac{c(A) - 0}{1 - 0} = c(A).$$

■

Lemma 11. $d^2 = d$

Proof: Since $m\big(d(A)\big) = 1$ by definition of (16), we have

$$d(d(A)) = \frac{A}{m(A)} = d(A).$$

■

Lemma 12. $(cd)^2 = cd$

Proof: Since $cd = dc$ (Theorem 8), $c^2 = c$ and $d^2 = d$, we have $cdcd = c^2d^2 = cd$. ■

Using Theorem 8 and Lemmas 10, 11 and 12, we can construct the multiplication table of the set $\{1, c, d, cd\}$.

Corollary 13. The multiplication table of $\{1, c, d, cd\}$ is given by

	1	c	d	cd
1	1	c	d	cd
c	c	c	cd	cd
d	d	cd	d	cd
cd	cd	cd	cd	cd

It follows from the above multiplication table that the set $\{1, c, d, cd\}$ is closed under multiplication (composition), is associative, and has an identity element. Hence, the set is a monoid. Since all elements commute and are idempotent, the set is a idempotent commutative monoid. In other words, $\{1, c, d, cd\}$ is a bounded semilattice. Note that the function cd acts as an absorbing element or zero element. Let $R = \mathbb{Z}\backslash 2\mathbb{Z}$ be the ring of integers modulo 2. The set $\{1, c, d, cd\}$ is isomorphic to R^2 with multiplication component wise, where1 $\mapsto$ (1,1), $c \mapsto$ (1,0), $d \mapsto$ (0,1) and $cd \mapsto$ (0,0).

Conclusions

In this paper, we have studied correction for chance and correction for maximum value as functions on a space of association coefficients. Various properties of both functions were presented. It was shown that the two functions commute under composition. Thus, if we want to correct a coefficient for chance and for maximum value, the result does not depend on the order in which the corrections are applied; and that if we correct for both chance and maximum value then a coefficient and all its linear transformations given the marginal totals are mapped to the same coefficient. In other words, if all linear transformations given the marginal totals of a particular coefficient that has zero value under independence are considered, then there is precisely one linear transformation that has maximum unity regardless of the marginal totals and zero value under independence. Finally, it was shown that the correction for chance function and the correction for maximum value function, together with the identity function and their composition, form a commutative idempotent monoid.

ACKNOWLEDGMENT

This research was funded by the Netherlands Organisation for Scientific Research, Veni project 451-11-026.

REFERENCES

Albatineh, A.N., and Niewiadomska-Bugaj, M. “MCS: A method for finding the number of clusters.'' Journal of Classification 28 (2011): 184-209.

Albatineh, A.N., Niewiadomska-Bugaj, M., and Mihalko, D. “On similarity indices and correction for chance agreement.'' Journal of Classification 23 (2006): 301-313.

Baulieu, F.B. “A classification of presence/absence based dissimilarity coefficients.'' Journal of Classification 6 (1989): 233-246.

Bennett, E.M., Alpert, R., and Goldstein, A.C. “Communications through limited response questioning.'' Public Opinion Quarterly 18 (1954): 303-308.

Bloch, D.A., and Kraemer, H.C. “2×2 Kappa coefficients: Measures of agreement or association.'' Biometrics 45 (1989): 269-287.

Brennan, R.L., and Light, R.J., “Measuring agreement when two observers classify people into categories not defined in advance.'' British Journal of Mathematical and Statistical Psychology 27 (1974): 154-163.

Brennan, R.L., and Prediger, D.J. “Coefficient kappa: Some uses, misuses, and alternatives.'' Educational and Psychological Measurement 41 (1981): 687-699.

Cohen, J. “A coefficient of agreement for nominal scales.'' Educational and Psychological Measurement 20 (1960): 213-220.

Davenport, E.C., and El-Sanhurry, N.A. “Phi/phimax: Review and synthesis.'' Educational and Psychological Measurement 51 (1991): 821-828.

Fleiss, J.L. “Measuring agreement between two judges on the presence or absence of a trait.'' Biometrics 31 (1975): 651-659.

Gower, J.C. “Some distance properties of latent root and vector methods used in multivariate analysis.'' Biometrika 53 (1966): 325-338.

Gower, J.C., and Legendre, P. “Metric and Euclidean properties of dissimilarity coefficients.'' Journal of Classification 3 (1986): 5-48.

Hanley, J.A. “Standard error of the kappa statistic.''

Psychological Bulletin 102 (1987): 315-321.

Hsu, L.M., and Field, R. “Interrater agreement measures: Comments on κ_n , Cohen's kappa, Scott's π and Aickin'sα." Understanding Statistics 2 (2003): 205-219.

Hubert, L.J. “Nominal scale response agreement as a generalized correlation." British Journal of Mathematical and Statistical Psychology 30 (1977): 98-103.

Hubert, L.J., and Arabie, P. “Comparing partitions." Journal of Classification 2 (1985): 193-218.

Janes, C.L. “An extension of the random error coefficient of agreement to *N*×*N* tables." British Journal of Psychiatry 134 (1979): 617-619.

Janson, S., and Vegelius, J. “On generalizations of the G index and the Phi coefficient to nominal scales." Multivariate Behavioral Research 14 (1979): 255-269.

Janson, S., and Vegelius, J. “The *J* index as a measure of nominal scale response agreement." Applied Psychological Measurement 6 (1982): 111-121.

Johnson, H.M. “Maximal selectivity, correctivity and correlation obtainable in a 2×2 contingency table." American Journal of Psychology 58 (1945): 65-68.

Krippendorff, K. “Association, agreement, and equity." Quality and Quantity 21 (1987): 109-123.

Krippendorff, K. “Reliability in content analysis. Some common misconceptions and recommendations." Human Communication Research 30 (2004a): 411-433.

Krippendorff, K. “Content Analysis: An Introduction to its Methodology." 2nd Edition. Sage, Thousands Oaks, CA, 2004b.

Loevinger, J.A. “A Systematic Approach to the Construction and Evaluation of Tests of Ability." Psychological Monographs: General and Applied, Vol 61, No 4, 1947.

Loevinger, J. “The technique of homogeneous tests compared with some aspects of ``scale analysis" and factor analysis." Psychological Bulletin 45 (1948): 507-529.

Maclure, M., and Willett, W.C. “Misinterpretation and misuse of the kappa statistic." American Journal of Epidemiology 126 (1987): 161-169.

Morey, L.C., and Agresti, A. “The measurement of classification agreement: An adjustment of the Rand statistic for chance agreement." Educational-and-PsychologicalMeasurement 44 (1984): 33-37.

Popping, R. “Overeenstemmingsmaten Voor Nominale Data." PhD thesis, Groningen, Rijksuniversiteit Groningen, 1983.

Rand, W.M. “Objective criteria for the evaluation of clustering methods." Journal of the American Statistical Association 66 (1971): 846-850.

Scott, W.A. “Reliability of content analysis: The case of nominal scale coding." Public Opinion Quarterly 19 (1955): 321-325.

Sijtsma, K., and Molenaar, I.W. “Introduction to Nonparametric Item Response Theory." Sage, Thousand Oaks, 2002.

Sokal, R.R., and Michener, C.D. “A statistical method for evaluating systematic relationships." University of Kansas Science Bulletin 38 (1958): 1409-1438.

Steinley, D. “Properties of the Hubert-Arabie adjusted Rand index." Psychological Methods 9 (2004): 386-396.

Warrens, M.J. “On the equivalence of Cohen's kappa and the Hubert-Arabie adjusted Rand index." Journal of Classification 25 (2008a): 177-183.

Warrens, M.J. “On association coefficients for 2×2 tables and properties that do not depend on the marginal distributions." Psychometrika 73 (2008b): 777-789.

Warrens, M.J. “On similarity coefficients for 2×2 tables and correction for chance." Psychometrika 73 (2008c): 487-502.

Warrens, M.J. “Inequalities between kappa and kappa-like statistics for *k*×*k* tables." Psychometrika 75 (2010): 176-185.

Warrens, M.J. “Chance-corrected measures for 2×2 tables that coincide with weighted kappa." British Journal of Mathematical and Statistical Psychology 64 (2011): 355-365.

Warrens, M.J. “The effect of combining categories on Bennett, Alpert and Goldstein's *S*." Statistical Methodology 9 (2012): 341-352.

Warrens, M.J. “On fixed points of the correction for chance function for 2×2 association coefficients. International Journal of Research and Reviews in Applied Sciences 15 (2013): 239-247.

Zegers, F.E. “A family of chance-corrected association coefficients for metric scales." Psychometrika 51 (1986a): 559-562.

Zegers, F.E. “A General Family of Association Coefficients." Groningen, Netherlands, Boomker, 1986b.

Zegers, F.E. “Coefficients for interrater agreement." Applied Psychological Measurement 15 (1991): 321-333.