PSYCHOMETRIKA—VOL. 73, NO. 4, 777–789
DECEMBER 2008
DOI: 10.1007/S11336-008-9070-3

ON ASSOCIATION COEFFICIENTS FOR 2×2 TABLES AND PROPERTIES THAT DO NOT DEPEND ON THE MARGINAL DISTRIBUTIONS

MATTHIJS J. WARRENS

LEIDEN UNIVERSITY INSTITUTE FOR PSYCHOLOGICAL RESEARCH

We discuss properties that association coefficients may have in general, e.g., zero value under statistical independence, and we examine coefficients for 2×2 tables with respect to these properties. Furthermore, we study a family of coefficients that are linear transformations of the observed proportion of agreement given the marginal probabilities. This family includes the phi coefficient and Cohen's kappa. The main result is that the linear transformations that set the value under independence at zero and the maximum value at unity, transform all coefficients in this family into the same underlying coefficient. This coefficient happens to be Loevinger's H.

Key words: agreement indices, resemblance measures, correction for chance, correction for maximum value, phi coefficient, Cohen's kappa, Hubert–Arabie adjusted Rand index, Yule's Q, Loevinger's H, Cole's C_7.

1. Introduction

Association coefficients are important tools in various domains of data analysis. Considerable literature is available concerning association coefficients for 2×2 tables (see, e.g., Janson & Vegelius, 1981; Gower & Legendre, 1986; Krippendorff, 1987; Hubálek, 1982; Baulieu, 1989, 1997; Albatineh, Niewiadomska-Bugaj & Mihalko, 2006). Well-known examples are the phi coefficient, Cohen's kappa, or the observed proportion of agreement, also known as the simple matching coefficient. Association coefficients for 2×2 tables are used, e.g., in biological ecology for measuring the degree of coexistence between two species types over different locations (cf. Sokal & Sneath, 1963), in psychology or biometrics for a 2×2 reliability study where two observers classify a sample of subjects using a dichotomous response (cf. Fleiss, 1975), or in cluster analysis for comparing two partitions of a set of objects obtained with different clustering algorithms (Albatineh et al., 2006; Steinley, 2004; Popping, 1983).

In this paper, we discuss several desiderata for association coefficients for 2×2 tables and several coefficients are examined with respect to these properties. Desiderata are properties that coefficients may have in general, not just for a particular set of data. The three properties that primarily concern us in this paper are zero value under statistical independence, maximum value unity, and minimum value minus unity independent of the marginal distributions. Coefficients for 2×2 tables that satisfy these three properties are the tetrachoric correlation, three transformations of the odds ratio, Yule's (1900) Q, Yule's (1912) Y, and Digby's (1983) H, and a measure of ecological association, Cole's (1949) C_7.

In addition, we study a family of coefficients that has been given a lot of attention in the literature. Coefficients that belong to this family are linear transformations of the observed proportion of agreement, given the marginal probabilities. The main result of the paper is that the linear transformations that set the value under independence at zero and the maximum value

Requests for reprints should be sent to Matthijs J. Warrens, Psychometrics and Research Methodology Group, Leiden University Institute for Psychological Research, Leiden University, Wassenaarseweg 52, P.O. Box 9555, 2300 RB Leiden, The Netherlands. E-mail: warrens@fsw.leidenuniv.nl

TABLE 1.
Bivariate proportions table for binary variables.

Variable one	Variable two		
	Value 1	Value 2	Total
Value 1	a	b	p_1
Value 2	c	d	q_1
Total	p_2	q_2	1

at unity, transform all coefficients in this family into the same underlying coefficient. This coefficient happens to be Loevinger's (1947, 1948) H. We conclude that if it is important that a coefficient has zero value under statistical independence, maximum value unity, and minimum value minus unity independent of the marginal distributions, then we may discard all coefficients that belong to the general family.

The paper is organized as follows. Definitions and examples of association coefficients for 2×2 tables are presented in the next section. In Sect. 3, we discuss desiderata for coefficients for 2×2 tables and several coefficients are examined with respect to the properties. Section 4 contains the main result. Section 5 contains the discussion.

2. Association Coefficients

Association coefficients are tools in data analysis that measure the strength of a relationship between two variables. A traditional measure for the 2×2 table is the tetrachoric correlation (Pearson, 1900; Divgi, 1979). The coefficient is an estimate of the Pearson product-moment correlation coefficient between hypothetical row and column variables with normal distributions that would reproduce the observed contingency table if they were divided into two categories in the appropriate proportions. Because an approximate estimate of the Pearson correlation may well be as adequate in many applications, particularly in small samples, various authors have introduced approximations to the tetrachoric correlation (Digby, 1983; Castellan, 1966; Pearson, 1900).

Many association coefficients for two binary variables can be defined using the four probabilities a, b, c, and d presented in Table 1. A counterexample is the tetrachoric correlation. The quantities a, b, c, and d characterize the joint distribution of the two variables. The row and column totals of Table 1 are the marginal distributions that result from summing the joint probabilities. We denote these by p_1 and q_1 for the first variable and by p_2 and q_2 for the second variable. Instead of probabilities, Table 1 may also be defined on counts or frequencies; probabilities are used here for notational convenience.

We will use S as a general symbol for a coefficient. Furthermore, following Sokal and Sneath (1963, p. 128) and Albatineh et al. (2006), the convention is adopted of calling a coefficient by its originator or the first we know to propose it. Moreover, we will study association coefficients as sample statistics and not as population parameters. In this section, we discuss several types of association coefficients for 2×2 tables, namely, transformations of the odds ratio, measures of ecological association, measures of interrater agreement, measures for comparing two partitions, and a measure for test homogeneity.

2.1. *Transformations of the Odds Ratio*

The odds ratio for Table 1 is defined as the ratio of the odds of an event occurring in one group (a/b) to the odds of it occurring in another group (c/d). These groups might be any other

dichotomous classification. An odds ratio of 1 indicates that the condition or event under study is equally likely in both groups. An odds ratio greater than 1 indicates that the event is more likely in the first group. Probability theory tells us that two binary variables are statistically independent if the odds ratio is equal to unity, i.e., $ad/bc = 1$. Due to the simple formula of its standard error, the logarithm of the odds ratio is sometimes preferred over the ordinary odds ratio. However, the value of both measures ranges from zero to plus infinity.

Edwards (1963) suggested that measures of association for 2×2 tables should be some function of the cross-product ad/bc. The coefficients

$$S_{\text{Yule2}} = \frac{\frac{ad}{bc} - 1}{\frac{ad}{bc} + 1} = \frac{ad - bc}{ad + bc} \quad \text{(Yule, 1900)},$$

$$S_{\text{Digby}} = \frac{(ad)^{3/4} - (bc)^{3/4}}{(ad)^{3/4} + (bc)^{3/4}} \quad \text{(Digby, 1983), and}$$

$$S_{\text{Yule3}} = \frac{(ad)^{1/2} - (bc)^{1/2}}{(ad)^{1/2} + (bc)^{1/2}} \quad \text{(Yule, 1912)}$$

transform the odds ratio to a correlation-like range $[-1, 1]$. Coefficients S_{Yule2}, S_{Digby}, and S_{Yule3} are nonlinear transformations of ad/bc.

Coefficient S_{Yule2} is Yule's coefficient of association, denoted by Q. Coefficient S_{Yule3} is Yule's coefficient of colligation, denoted by Y. Coefficients S_{Yule2}, S_{Digby}, and S_{Yule3} have been used as approximations to the tetrachoric correlation. Some properties of S_{Yule2} and S_{Yule3} are discussed in Castellan (1966). Coefficient S_{Digby} has a value between S_{Yule2} and S_{Yule3}. Digby (1983) uses the symbol H for S_{Digby} and shows that the coefficient performs better than S_{Yule2} and S_{Yule3} as an approximation to the tetrachoric correlation.

2.2. *Ecological Association*

In ecological biology, one may distinguish several contexts where association coefficients have been used (Janson & Vegelius, 1981). One such case deals with measuring the degree of coexistence between two species over different locations. A second situation is measuring association between two locations over different species. In the first situation, a binary variable is a coding of the presence or absence of a species in a number of locations. The joint probability a then equals the proportion of locations that two species have in common.

Janson and Vegelius (1981) require that measures of ecological association satisfy the following three properties. A measure of coexistence should not be based on d, which in ecology is the proportion of mismatches. Basing similarity between two species on the mutual absence of a certain character is considered improper (cf. Sokal & Sneath, 1963). The second requirement is that the minimum value of S satisfies the condition $S = 0 \Leftrightarrow a = 0$. Since probability a denotes the proportion of locations where two species types both exist, the minimum value should be taken if and only if two species types are never found together. The third requirement is that the maximum value of S satisfies $S = 1 \Leftrightarrow b = c = 0$. The maximum coexistence must occur when two species types always occur together. Association coefficients that satisfy the three properties are

$$S_{\text{Jac}} = \frac{a}{a + b + c} = \frac{a}{p_1 + p_2 - a} \quad \text{(Jaccard, 1912)},$$

$$S_{\text{Dice}} = \frac{2a}{2a + b + c} = \frac{2a}{p_1 + p_2} \quad \text{(Dice, 1945), and}$$

$$S_{\text{Och}} = \frac{a}{\sqrt{p_1 p_2}} \quad \text{(Ochiai, 1957)}.$$

Coefficient S_{Jac} may be interpreted as the proportion of locations where two species types both exist, divided by the proportion of locations where at least one of the species types exists. Coefficients S_{Dice} and S_{Och} are the harmonic and geometric mean of the conditional probabilities a/p_1 and a/p_2 (Janson & Vegelius, 1981). Coefficient S_{Dice} gives twice as much weight to a compared to S_{Jac}, and is used if a is relatively small compared to b and c.

Coefficients S_{Jac}, S_{Dice}, and S_{Och} are popular measures of ecological association, and they have been empirically compared to other coefficients for 2×2 tables in numerous studies. For example, Duarte, Santos, and Melo (1999) evaluated association measures in clustering and ordination of common bean cultivars analyzed by RAPD type molecular markers. The genetic distance measures obtained by taking the complement of the Dice coefficient were considered the most adequate. Boyce and Ellison (2001) studied similarity coefficients for 2×2 tables in the context of fuzzy set ordination, and concluded that the Dice, Ochiai, and Jaccard coefficients are the preferred association measures.

A fourth measure of ecological association is

$$S_{\text{Sim}} = \frac{a}{\min(p_1, p_2)} \quad \text{(Simpson, 1943)}.$$

Coefficient S_{Sim} is the maximum of the conditional probabilities a/p_1 and a/p_2. The coefficient is very similar to S_{Jac}, S_{Dice}, and S_{Och}. The difference is that S_{Sim} obtains its maximum value of unity if the two species types have a deterministic relationship. Coefficient $S_{\text{Sim}} = 1$ if one species type only occurs in locations where the second type exists. The second type may occur in places where the first type is not found.

The above four coefficients measure the degree to which two species types occur jointly in a number of locations. Several authors proposed coefficients of ecological association that measure the degree to which the observed proportion of joint occurrences of two species types exceeds or falls short of the proportion of joint occurrences expected on the basis of chance alone (cf. Cole, 1949). A measure introduced in Cole (1949) can be written as

$$S_{\text{Cole}} = \begin{cases} (ad - bc)/\min(p_1 q_2, p_2 q_1) & \text{if } ad > bc, \\ 0 & \text{if } ad = bc, \\ (ad - bc)/\min(p_1 p_2, q_1 q_2) & \text{if } ad < bc. \end{cases}$$

This (correct) formula, S_{Cole}, can be found in Ratliff (1982). Coefficient S_{Cole}, also denoted as C_7, is equivalent to Loevinger's (1947, 1948) H (Sect. 2.5) if $ad \geq bc$, i.e., if the two binary variables are positively dependent.

Although S_{Cole} is less popular than measures S_{Jac}, S_{Dice}, and S_{Och}, the coefficient has been used in various applications by animal and plant ecologists (Hurlbert, 1969; Ratliff, 1982). A variant of S_{Cole}, C_8 proposed in Hurlbert (1969), is less influenced by the species' frequencies. Hurlbert (1969) examined both C_7 and C_8 as approximations to the tetrachoric correlation.

2.3. Interrater Agreement

Suppose the variables are observers and that Table 1 is the cross classification of the judgments by the two raters on the presence or absence of a trait. An obvious measure of agreement that has been proposed independently for this situation by various authors (Fleiss, 1975), is the proportion of all subjects on whom the two raters agree, $a + d$ (see, e.g., Goodman & Kruskal, 1954). The observed proportion of agreement, $S_{\text{SM}} = a + d$, is also referred to as the simple matching coefficient (Sokal & Michener, 1958). In this domain of data analysis, it is considered a necessity that the agreement measure is chance corrected. Examples of chance-corrected

agreement indices are

$$S_{\text{Yule1}} = \frac{ad - bc}{\sqrt{p_1 p_2 q_1 q_2}} \quad \text{(Yule, 1912)},$$

$$S_{\text{Cohen}} = \frac{2(ad - bc)}{p_1 q_2 + p_2 q_1} \quad \text{(Cohen, 1960), and}$$

$$S_{\text{MP}} = \frac{2(ad - bc)}{p_1 q_1 + p_2 q_2} \quad \text{(Maxwell \& Pilliner, 1968)}.$$

The phi coefficient S_{Yule1} is what the Pearson product-moment correlation becomes when it is applied to binary variables. This correlation plays an important role in educational and psychological measurement. Coefficient S_{Cohen} is Cohen's kappa for two binary variables, a popular measure for interrater agreement. Although coefficients S_{Yule1}, S_{Cohen}, and S_{MP} have a correlation-like range $[-1, 1]$, the coefficients are usually used to distinguish between positive and zero association.

2.4. *Comparing Two Partitions*

In cluster analysis, one may be interested in comparing the partitions from two different clustering methods (Rand, 1971; Hubert & Arabie, 1985; Steinley, 2004; Albatineh et al., 2006; Popping, 1983). An equivalent problem in psychology is that of measuring agreement among judges in classifying answers to open-ended questions, or psychologists rating people on categories not defined in advance (Brennan & Light, 1974; Janson & Vegelius, 1982; Popping, 1983, 1984). Suppose we have two partitions of the same objects. The two clustering partitions can be summarized by a 2×2 table with quantities a, b, c, and d, by counting the number of pairs of objects that were placed in the same cluster in both partitions (a), in the same cluster in one partition but in different clusters in the other partition (b and c), and in different clusters in both (d). Next, one may use an association measure for a 2×2 table that quantifies the amount of agreement between the two partitions.

For some time, the Rand index

$$S_{\text{Rand}} = \frac{a + d}{a + b + c + d} \quad \text{(Rand, 1971)}$$

was a popular measure for comparing two partitions. Coefficient S_{Rand} is equivalent to the simple matching coefficient and the measure proposed in Brennan and Light (1974) for measuring agreement among psychologists rating people on categories not defined in advance. Nowadays, there seems to be some agreement in the cluster community that the preferred measure for comparing two partitions is the Hubert–Arabie (1985) adjusted Rand index (cf. Steinley, 2004). Warrens (in press) shows that the Hubert–Arabie adjusted Rand index can be written as

$$S_{\text{HA}} = \frac{2(ad - bc)}{p_1 q_2 + p_2 q_1} \quad \text{(Hubert \& Arabie, 1985)}.$$

The adjusted Rand index S_{HA} is thus equivalent to Cohen's kappa for two categories.

2.5. *Test Homogeneity*

Consider the Loevinger (1947, 1948) coefficient

$$S_{\text{Loe}} = \frac{ad - bc}{\min(p_1 q_2, p_2 q_1)}$$

also denoted by H. Coefficient S_{Loe} is a central 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 (cf. Sijtsma & Molenaar, 2002). Coefficient S_{Loe} is attributed to Loevinger (1947, 1948) by Mokken (1971) and Sijtsma and Molenaar (2002). Krippendorff (1987) reports that S_{Loe} is already discussed in Benini (1901).

Although coefficient S_{Loe} has a correlation-like range $[-1, 1]$, it is usual to assume that two items are at least positively dependent. If two binary variables have positive covariance, coefficient S_{Loe} is equivalent to the measure of ecological association S_{Cole} (Sect. 2.2). Coefficient $S_{\text{Loe}} = 1$ if two items form a so-called Guttman pair, i.e., if $a = \min(p_1, p_2)$. In this case, all subjects that pass the first item also pass the second item or vice versa. Using S_{Loe}, we may have perfect association with different marginal distributions, i.e., the item popularities or difficulties p_1 and p_2 may be different.

3. Desiderata

Recall that desiderata are properties that coefficients may have in general, not just for a particular set of data. Several authors, among whom Baulieu (1989, 1997), Zegers (1986), Popping (1983), and Janson and Vegelius (1981), have formulated desiderata for association coefficients. A basic requirement is that S is symmetric. An association coefficient is symmetric if it is insensitive to the order of the variables. Asymmetric coefficients may be encountered in situations where one of the variables can be regarded as a standard. A second basic requirement is that the association of a variable with itself is perfect. The association between two variables is perfect if a coefficient attains its maximum value. The maximum value should be obtained if a variable is compared with itself.

In this paper, we are particularly interested in desiderata that deal with the maximum value, zero value, and minimum value of a coefficient. Depending on the domain of data analysis, different properties may or may not be considered desirable.

3.1. Zero Value under Statistical Independence

The following property is also discussed in Zegers (1986) and Popping (1983).

(d1) $S = 0$ if two variables are statistically independent.

In several domains of data analysis, (d1) is a natural desideratum. Property (d1) is not particularly important for coefficients of ecological association discussed in Sect. 2.2 that measure the degree to which two species occur jointly. However, chance-corrected coefficients, e.g., coefficient S_{Cole}, have been introduced in biological ecology. Furthermore, property (d1) is considered a necessity for reliability studies in which two observers judge each a sample of subjects on a binary trait. A popular chance-corrected measure for this type of data is Cohen's kappa (the Hubert–Arabie adjusted Rand index S_{HA} is thus also a chance-corrected measure). A fourth coefficient that satisfies (d1) is the Pearson product-moment correlation coefficient (with special case S_{Yule1}).

The expected values of probabilities a, b, c, and d given the marginal distributions, i.e., the values under statistical independence, are presented in Table 2. The expected value of a, denoted by $E(a)$, can be obtained by considering all permutations of the observations of one of the two variables, while preserving the order of the observations of the other variable. For each permutation, the value of a can be determined. The mean of these values is $p_1 p_2$. The expectation of each cell of the 2×2 table is equal to the product of the corresponding marginal probabilities.

TABLE 2.
Expected values of a, b, c, and d from Table 1 under statistical independence given fixed marginal distributions.

Variable one	Variable two		
	Value 1	Value 2	Total
Value 1	p_1p_2	p_1q_2	p_1
Value 2	p_2q_1	q_1q_2	q_1
Total	p_2	q_2	1

If a coefficient does not satisfy requirement (d1), it can be corrected for association due to chance using the linear transformation

$$CS = \frac{S - E(S)}{\max(S) - E(S)}. \tag{1}$$

In correction (1), CS is the symbol for a corrected coefficient, $E(S)$ denotes the expected value of S under statistical independence, and $\max(S)$ is the maximum value of S regardless of the marginal probabilities. For all coefficients considered in this paper, $\max(S) = 1$.

Both Fleiss (1975) and Zegers (1986) showed that S_{Cohen} may be interpreted as a chance-corrected version of S_{SM} and S_{Dice}.

Proposition 1. *Coefficients S_{SM} and S_{Dice} become S_{Cohen} after correction (1).*

Proof: We consider the proof for S_{SM} only. The proof for S_{Dice} is similar.

$E(S_{\text{SM}}) = E(a + d) = p_1p_2 + q_1q_2$. Using S_{SM}, $E(S_{\text{SM}})$ and $\max(S_{\text{SM}}) = 1$ in (1), we obtain

$$CS_{\text{SM}} = \frac{a + d - p_1p_2 - q_1q_2}{1 - p_1p_2 - q_1q_2}. \tag{2}$$

We have the identities

$$a - p_1p_2 = a - (a + b)(a + c) = a(1 - a - b - c) - bc = ad - bc, \tag{3}$$

$$d - q_1q_2 = d - (b + d)(c + d) = d(1 - b - c - d) - bc = ad - bc, \tag{4}$$

and

$$1 = (p_1 + q_1)(p_2 + q_2) = p_1p_2 + p_1q_2 + p_2q_1 + q_1q_2. \tag{5}$$

Using (3) and (4) in the numerator of (2), and using (5) in the denominator of (2), yields

$$CS_{\text{SM}} = \frac{2(ad - bc)}{p_1q_2 + p_2q_1} = S_{\text{Cohen}}. \qquad \square$$

3.2. *Maximum Value Independent of Marginal Probabilities*

In general, we speak of perfect positive association between two variables if $S = 1$. However, we may require the following stronger property for an association coefficient.

(d2) Maximum $S = 1$ independent of the marginal distributions.

Coefficients that satisfy requirement (d2) are S_{Yule2}, S_{Digby}, and S_{Yule3} (Sect. 2.1), S_{Sim} and S_{Cole} (Sect. 2.2), and S_{Loe} (Sect. 2.5).

Suppose the binary variables are two test items that subjects may either pass or fail, and suppose we want to study the association between these items. In a test-theoretical application, one

may want a coefficient that is 1 if two items have a deterministic relationship, i.e., passing the first item implies passing the second item. If all subjects that pass the first (more difficult) item also pass the second (easier) item, we speak of a so-called Guttman pair. Using an association coefficient that satisfies (d2), we may have perfect association with different marginal distributions, i.e., the item popularities or difficulties p_1 and p_2 may be different. As a second example, suppose that the binary variables are the codings of the presence or absence of two species in a number of locations. A biologist may want a coefficient that is 1 if the occurrence of one species type also implies the existence of another species type for a location. The second species type may nevertheless occur in places where the first type is not found.

In several domains of data analysis, (d2) is not a natural requirement. Consider a reliability study in which two observers judge each a sample of subjects on a binary trait. A popular measure for interrater agreement is Cohen's kappa. If the marginal distributions are not the same, i.e., if there exists marginal asymmetry, the two raters do not agree, and it may be argued that a coefficient of agreement like S_{Cohen} should never be unity in this case. In other words, if different marginal distributions reflect disagreement, it may be desirable that the value of the coefficient is smaller than unity.

For many association coefficients, the maximal attainable value depends on the marginal distributions. For example, probability a in Table 1 cannot exceed its marginal probabilities p_1 and p_2. The measures of association in Sect. 2.2, e.g., S_{Jac} or S_{Dice}, can therefore only attain the maximum value of unity if $p_1 = p_2$, i.e., in the case of marginal symmetry (coefficient S_{Sim} and S_{Cole} are exceptions). The maximum value of a, denoted by $a_{\max}$, equals $a_{\max} = \min(p_1, p_2)$. The maximum value of S_{Dice} given the marginal distributions equals $2\min(p_1, p_2)/(p_1 + p_2)$.

The maximum value of the covariance $(ad - bc)$ between two binary variables given the marginal distributions is equal to $(ad - bc)_{\max} = \min(p_1q_2, p_2q_1)$. The maximum value of the phi coefficient S_{Yule1} (and other coefficients from Sect. 2.3) is thus also restricted by the marginal distributions (Cureton, 1959; Guilford, 1965; Zysno, 1997). In the literature on this phenomenon, it was suggested to use the ratio S_{Yule1} divided by the maximum value of S_{Yule1} given the marginal probabilities. A detailed review of the phi/phimax literature is presented in Davenport and El-Sanhurry (1991). In general, for coefficients of which the maximum value depends on the marginal probabilities, authors from the phi/phimax literature suggest the linear transformation

$$MS = \frac{S}{S_{\max}}. \tag{6}$$

In correction (6), MS is the symbol for a corrected coefficient and $S_{\max}$ is the maximum value of S given the marginal distributions.

Proposition 2 combines a result in Loevinger (1948, p. 519) and a result in Davenport and El-Sanhurry (1991).

Proposition 2. *Division of* S_{Yule1}, S_{Cohen}, *and* S_{MP} *by their maximum values given the marginal distributions*, *yields* S_{Loe}.

Proof: We only consider the proof for S_{Yule1}. The proof for the other coefficients is similar.

The maximum value of S_{Yule1} given the marginal distributions is

$$\frac{\min(p_1q_2, p_2q_1)}{\sqrt{p_1p_2q_1q_2}}. \tag{7}$$

Using S_{Yule1} and (7) in (6), we obtain S_{Loe}. □

Proposition 3. *Division of* S_{Dice} *and* S_{Och} *by their maximum values given the marginal distributions*, *yields* S_{Sim}.

Proof: We only consider the proof for S_{Dice}. The proof for S_{Och} is similar.

The maximum value of S_{Dice} given the marginal distributions is equal to $2\min(p_1, p_2)/(p_1 + p_2)$. Using S_{Dice} and $2\min(p_1, p_2)/(p_1 + p_2)$ in (6), we obtain S_{Sim}. □

Thus, coefficient S_{Loe} may be interpreted as a normed version of S_{Yule1}, S_{Cohen}, or S_{MP}. Furthermore, coefficient S_{Sim} may be interpreted as a normed version of S_{Dice} and S_{Och}.

3.3. Minimum Value Independent of Marginal Probabilities

Instead of distinguishing between positive and zero association, it may be required to distinguish between positive, zero, and negative association. In general, we speak of perfect negative association between two variables if $S = -1$. However, we may require the following stronger property for an association coefficient.

(d3) Minimum $S = -1$ independent of the marginal distributions.

The following situation was pointed out by one of the reviewers. Consider workers in nursing homes and let the binary variables be codings of nursing and of caring. Some people nurse, some people care, some do both, and some do neither nursing nor caring. Suppose we want to study the association between nursing and caring. We may require a coefficient that is 1 if nursing and caring have a deterministic positive relation (nursing implies caring, or caring implies nursing), -1 if they have a deterministic negative relation (nursing does not imply caring or caring does not imply nursing), and 0 if nursing and caring are independent.

Coefficients that satisfy requirement (d3) are the tetrachoric correlation, the three transformations of the odds ratio discussed in Sect. 2.1, S_{Yule2}, S_{Digby}, and S_{Yule3}, and the measure of ecological association discussed in Sect. 2.2, S_{Cole}. Moreover, these coefficients are the only coefficients discussed in this paper that satisfy the three requirements (d1), (d2), and (d3) jointly.

4. Linear Transformations of S_{SM}

The coefficients in Sects. 2.2 to 2.5 come from different domains of data analysis. Although these measures are applied in different contexts, the formulas of the different coefficients are related. Consider a family $\mathcal{L}$ of coefficients of a form $\lambda + \mu(a + d)$, where probabilities a and d are defined in Table 1, and where λ and μ, different for each coefficient, depend on the marginal distributions. Since the observed proportion of agreement is defined as $S_{\text{SM}} = a + d$, coefficients in the $\mathcal{L}$ family are linear transformations of S_{SM}, the simple matching coefficient, given the marginal probabilities.

The $\mathcal{L}$ family has been given a lot of attention in the literature. Clearly, S_{SM} $(=S_{\text{Rand}})$ is in the $\mathcal{L}$ family. Other coefficients that belong to $\mathcal{L}$ are S_{Yule1}, S_{Cohen} $(=S_{\text{HA}})$, and S_{MP} (Sect. 2.3), and S_{Loe} (Sect. 2.5). For example, using identities (3) and (4), coefficient S_{Cohen} can be written as $S_{\text{Cohen}} = \lambda + \mu(a + d)$ where

$$\lambda = -\frac{p_1 p_2 + q_1 q_2}{p_1 q_2 + p_2 q_1} \quad \text{and} \quad \mu = \frac{1}{p_1 q_2 + p_2 q_1}.$$

Furthermore, since $a = p_2 - q_1 + d$, probabilities a and d are also linear in $(a + d)$. Linear in $(a + d)$ is therefore equivalent to linear in a and linear in d.

Proposition 4. *Coefficients of a form $T = \kappa + \nu a$, where κ and ν, different for each coefficient, depend on the marginal distributions, are in the $\mathcal{L}$ family.*

Proof: We have $a = p_2 - q_1 + d$. Hence, T can be written as $\lambda + \mu(a + d)$ where

$$\lambda = \frac{2\kappa + \nu(p_2 - q_1)}{2} \quad \text{and} \quad \mu = \frac{\nu}{2}. \qquad \square$$

Coefficients that are linear in a are S_{Dice}, S_{Och}, and S_{Sim} (Sect. 2.2). Coefficients that are not in the $\mathcal{L}$ family are coefficient S_{Jac} (Sect. 2.2) and the transformations of the odds ratio discussed in Sect. 2.1. Furthermore, coefficient S_{Cole} is defined by two different linear transformations of $(a + d)$, one for $ad < bc$ and one for $ad > bc$. Coefficient S_{Cole} is therefore not a member of the $\mathcal{L}$ family.

In Sects. 2 and 3 we discussed association coefficients for 2×2 tables and properties that these coefficients may satisfy in general. The only coefficients that have zero value under statistical independence, maximum value unity, and minimum value minus unity independent of the marginal distributions, are the transformations of the odds ratio, S_{Yule2}, S_{Digby}, and S_{Yule3}, discussed in Sect. 2.1, and a measure of ecological association, S_{Cole}. As it turns out, none of the other coefficients discussed in this paper satisfy the three requirements (d1), (d2), and (d3) jointly. More precisely, there is no coefficient in the $\mathcal{L}$ family (coefficients of a form $\lambda + \mu(a + d)$) that satisfies (d1), (d2), and (d3) jointly. Furthermore, there is exactly one member in $\mathcal{L}$ that satisfies (d1) and (d2).

The theorem below shows that the two linear transformations (1) and (6) that set the value under independence at zero and the maximum value at unity, transform all coefficients in $\mathcal{L}$ into the same underlying coefficient. This coefficient happens to be S_{Loe}. For notational convenience, we provide the proof for coefficients that are linear transformations of joint probability a given the marginal distributions.

Theorem 1. *A coefficient of a form $\kappa + \nu a$ becomes coefficient S_{Loe} after corrections* (1) *and* (6)*, irrespective of the order of the transformations.*

Proof: Using $\kappa + \nu a$ and its maximum value given the marginal distributions, $\kappa + \nu \min(p_1, p_2)$, in (6), we obtain

$$\frac{\kappa + \nu a}{\kappa + \nu \min(p_1, p_2)}. \tag{8}$$

The expected value of (8) given the marginal distributions is equal to

$$E\left(\frac{\kappa + \nu a}{\kappa + \nu \min(p_1, p_2)}\right) = \frac{\kappa + \nu E(a)}{\kappa + \nu \min(p_1, p_2)} = \frac{\kappa + \nu p_1 p_2}{\kappa + \nu \min(p_1, p_2)}. \tag{9}$$

The maximal value of (8) regardless of the marginal probabilities is equal to

$$\max\left(\frac{\kappa + \nu a}{\kappa + \nu \min(p_1, p_2)}\right) = \frac{\kappa + \nu \min(p_1, p_2)}{\kappa + \nu \min(p_1, p_2)} = 1. \tag{10}$$

Using (8), its expectation (9), and (10) in (1), and multiplying the result by $\kappa + \nu \min(p_1, p_2)$, we obtain

$$\frac{\kappa + \nu a - \kappa - \nu p_1 p_2}{\kappa + \nu \min(p_1, p_2) - \kappa - \nu p_1 p_2} = \frac{a - p_1 p_2}{\min(p_1, p_2) - p_1 p_2} = S_{\text{Loe}}.$$

Alternatively, using $\kappa + \nu a$ and its expected value $\kappa + \nu p_1 p_2$ in (1), we obtain

$$\frac{\kappa + \nu a - \kappa - \nu p_1 p_2}{\max(\kappa + \nu a) - \kappa - \nu p_1 p_2} = \frac{a - p_1 p_2}{[\max(\kappa + \nu a) - \kappa]/\nu - p_1 p_2}. \tag{11}$$

The maximum value of (11) given the marginal distributions is equal to

$$\frac{\min(p_1, p_2) - p_1 p_2}{[\max(\kappa + \nu a) - \kappa]/\nu - p_1 p_2}. \tag{12}$$

Using (11) and its maximum value (12) given the marginal distributions in (6), we obtain

$$\frac{a - p_1 p_2}{\min(p_1, p_2) - p_1 p_2} = S_{\text{Loe}}.$$

This completes the proof. □

5. Discussion

In this paper, we discussed association coefficients for 2×2 tables and properties that these coefficients may satisfy in general. Coefficients that have zero value under statistical independence, maximum value unity, and minimum value minus unity independent of the marginal distributions, are the tetrachoric correlation, three transformations of the odds ratio, Yule's (1900) Q, Yule's (1912) Y, and Digby's (1983) H, and a measure of ecological association, Cole's (1949) C_7. The latter four coefficients have been studied as approximations to the tetrachoric correlation. Yule's Q and Yule's Y (together with the Jaccard, Dice, and Ochiai coefficients discussed in Sect. 2.2, and the simple matching and phi coefficients discussed in Sect. 2.3) are implemented in the hierarchical cluster routine of the software package SPSS 14.0.

For a general family $\mathcal{L}$ of coefficients that are linear transformations of the observed proportion of agreement, it was shown that the two linear transformations that set the value under independence at zero and the maximum value at unity transform all coefficients in $\mathcal{L}$ family into the same underlying coefficient. This coefficient is Loevinger's H (Loevinger, 1947, 1948). Loevinger's H and Cole's C_7 are equivalent if the binary variables are positively dependent. Cole's C_7 has zero value under statistical independence, maximum value unity, and minimum value minus unity independent of the marginal distributions, but is not in the $\mathcal{L}$ family. If all three desiderata are important, then we may discard all coefficients that belong to the $\mathcal{L}$ family. The fact that no coefficient in $\mathcal{L}$ satisfies all three properties might explain why there is an almost endless list of 2×2 coefficients. Furthermore, if it is important that a coefficient has zero value under statistical independence and maximum value unity independent of the marginal distributions, then we may discard all coefficients that belong to the $\mathcal{L}$ family, except Loevinger's H.

Loevinger's H is the only linear transformation of the observed proportion of agreement that has zero value under independence and maximum unity independent of the marginal distributions. Because its minimum value is not -1 with different marginal distributions, Loevinger's H is more directed to positive association than to negative association. The coefficient is a logical choice in cases where positive association needs to be distinguished from zero association, e.g., analyzing test items. If both positive and negative association are important, Yule's Q and Yule's Y are a logical choice. Consider the case that two raters each judge a number of people on the presence or absence of a trait. For this situation it is usual to distinguish between positive and zero association. Furthermore, if the marginal distributions are not the same, i.e., if there exists marginal asymmetry, the two raters do not agree perfectly, and it is undesirable that the value of the coefficient of agreement is unity. Loevinger's H is therefore not a logical choice for measuring interrater agreement. A popular measure for interrater agreement is Cohen's (1960) kappa. For the case of two categories, Cohen's kappa is equivalent to the Hubert–Arabie (1985) adjusted Rand index (Warrens, in press).

References

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

Baulieu, F.B. (1989). A classification of presence/absence based dissimilarity coefficients. *Journal of Classification*, *6*, 233–246.

Baulieu, F.B. (1997). Two variant axiom systems for presence/absence based dissimilarity coefficients. *Journal of Classification*, *14*, 159–170.

Benini, R. (1901). *Principii di demografie. No. 29 of manuali Barbèra di science giuridiche sociali e poltiche*. Firenzi: G. Barbèra.

Boyce, R.L., & Ellison, P.C. (2001). Choosing the best similarity index when performing fuzzy set ordination on binary data. *Journal of Vegetational Science*, *12*, 711–720.

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

Castellan, N.J. (1966). On the estimation of the tetrachoric correlation coefficient. *Psychometrika*, *31*, 67–73.

Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, *20*, 37–46.

Cole, L.C. (1949). The measurement of interspecific association. *Ecology*, *30*, 411–424.

Cureton, E.E. (1959). Note on $\phi/\phi_{\max}$. *Psychometrika*, *24*, 89–91.

Davenport, E.C., & El-Sanhurry, N.A. (1991). Phi/phimax: review and synthesis. *Educational and Psychological Measurement*, *51*, 821–828.

Dice, L.R. (1945). Measures of the amount of ecologic association between species. *Ecology*, *26*, 297–302.

Digby, P.G.N. (1983). Approximating the tetrachoric correlation coefficient. *Biometrics*, *39*, 753–757.

Divgi, D.R. (1979). Calculation of the tetrachoric correlation coefficient. *Psychometrika*, *44*, 169–172.

Duarte, J.M., Santos, J.B., & Melo, L.C. (1999). Comparison of similarity coefficients based on RAPD markers in the common bean. *Genetics and Molecular Biology*, *22*, 427–432.

Edwards, A.W.F. (1963). The measure of association in a 2×2 table. *Journal of the Royal Statistical Society, Series A*, *126*, 109–114.

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

Goodman, L.A., & Kruskal, W.H. (1954). Measures of association for cross classifications. *Journal of the American Statistical Association*, *49*, 732–764.

Gower, J.C., & Legendre, P. (1986). Metric and Euclidean properties of dissimilarity coefficients. *Journal of Classification*, *3*, 5–48.

Guilford, J.P. (1965). The minimal phi coefficient and the maximal phi. *Educational and Psychological Measurement*, *25*, 3–8.

Hubálek, Z. (1982). Coefficients of association and similarity based on binary (presence-absence) data: an evaluation. *Biological Reviews*, *57*, 669–689.

Hubert, L.J., & Arabie, P. (1985). Comparing partitions. *Journal of Classification*, *2*, 193–218.

Hurlbert, S.H. (1969). A coefficient of interspecific association. *Ecology*, *50*, 1–9.

Jaccard, P. (1912). The distribution of the flora in the Alpine zone. *The New Phytologist*, *11*, 37–50.

Janson, S., & Vegelius, J. (1981). Measures of ecological association. *Oecologia*, *49*, 371–376.

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

Krippendorff, K. (1987). Association, agreement, and equity. *Quality and Quantity*, *21*, 109–123.

Loevinger, J.A. (1947). A systematic approach to the construction and evaluation of tests of ability. *Psychometrika Monograph No. 4*.

Loevinger, J.A. (1948). The technique of homogeneous tests compared with some aspects of scale analysis and factor analysis. *Psychological Bulletin*, *45*, 507–530.

Maxwell, A.E., & Pilliner, A.E.G. (1968). Deriving coefficients of reliability and agreement for ratings. *British Journal of Mathematical and Statistical Psychology*, *21*, 105–116.

Mokken, R.J. (1971). *A theory and procedure of scale analysis*. Hague: Mouton.

Ochiai, A. (1957). Zoogeographic studies on the soleoid fishes found in Japan and its neighboring regions. *Bulletin of the Japanese Society for Fish Science*, *22*, 526–530.

Pearson, K. (1900). Mathematical contributions to the theory of evolution. VII. On the correlation of characters not quantitatively measurable. *Philosophical Transactions of the Royal Society of London, Series A*, *195*, 1–47.

Popping, R. (1983). *Overeenstemmingsmaten Voor Nominale Data*. Unpublished doctoral dissertation, Rijksuniversiteit Groningen, Groningen, The Netherlands.

Popping, R. (1984). Traces of agreement. On some agreement indices for open-ended questions. *Quality and Quantity*, *18*, 147–158.

Rand, W. (1971). Objective criteria for the evaluation of clustering methods. *Journal of the American Statistical Association*, *66*, 846–850.

Ratliff, R.D. (1982). A correction of Cole's C_7 and Hurlbert's C_8 coefficients of interspecific association. *Ecology*, *50*, 1–9.
Sijtsma, K., & Molenaar, I.W. (2002). *Introduction to nonparametric item response theory*. Thousand Oaks: Sage.
Simpson, G.G. (1943). Mammals and the nature of continents. *American Journal of Science*, *24*, 11–31.
Sokal, R.R., & Michener, C.D. (1958). A statistical method for evaluating systematic relationships. *University of Kansas Science Bulletin*, *38*, 1409–1438.
Sokal, R.R., & Sneath, P.H. (1963). *Principles of numerical taxonomy*. San Francisco: Freeman.
Steinley, D. (2004). Properties of the Hubert–Arabie adjusted Rand index. *Psychological Methods*, *9*, 386–396.
Warrens, M.J. (in press). On the equivalence of Cohen's kappa and the Hubert–Arabie adjusted Rand index. *Journal of Classification*.
Yule, G.U. (1900). On the association of attributes in statistics. *Philosophical Transactions of the Royal Society A*, *75*, 257–319.
Yule, G.U. (1912). On the methods of measuring the association between two attributes. *Journal of the Royal Statistical Society*, *75*, 579–652.
Zegers, F.E. (1986). *A general family of association coefficients*. Unpublished doctoral dissertation, Rijksuniversiteit Groningen, Groningen, The Netherlands.
Zysno, P.V. (1997). The modification of the phi-coefficient reducing its dependence on the marginal distributions. *Methods of Psychological Research Online*, *2*, 41–52.

Manuscript Received: 7 DEC 2007
Final Version Received: 15 MAY 2008
Published Online Date: 23 JUL 2008