Correction for chance and correction for maximum value

Matthijs J Warrens

IFCS Bologna
July 6, 2015
Introduction

Association coefficients
- Quantify the degree of association between two variables
- Important tools in data analysis and classification

Desirable properties (in some contexts)
- Value 0 under statistical independence
- Maximum value 1 (regardless of marginal distributions)

Transformations
- Value 0: correction for chance
- Value 1: correction for maximum value

Both transformations can be studied as mathematical functions
Contingency tables

Contingency table \( \{ p_{ij} \} \) of size \( k \times \ell \) where \( k, \ell \geq 2 \)

Marginal totals

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

The set

\[
M = \left\{ \{ p_{ij} \}_{k \times \ell} \mid p_{ij} \geq 0 \text{ for all } i, j; \sum_{i,j} p_{ij} = 1 \right\}
\]

is the domain of the coefficients considered here
Examples for $2 \times 2$ tables

Phi coefficient

$$\phi = \frac{p_{11}p_{22} - p_{12}p_{21}}{\sqrt{p_1+p_2+p_1p_2}}$$

- value 0 under statistical independence

Loevinger’s $H$

$$H = \frac{p_{11}p_{22} - p_{12}p_{21}}{\min \{p_1+p_2, p_1p_2\}}$$

- an important statistic in Mokken scale analysis
- value 0 under statistical independence
- value 1 always attainable
Examples for $k \times k$ tables

Categorical variables with identical categories

Overall agreement $O = \sum_{i=1}^{k} p_{ii}$
- value 0 if no agreement
- value 1 if perfect agreement

Cohen’s kappa

$$\kappa = \frac{\sum_{i=1}^{k} (p_{ii} - p_{i+}p_{+i})}{1 - \sum_{i=1}^{k} p_{i+}p_{+i}}$$

- value 0 under statistical independence
Domain \( M = \left\{ \{p_{ij}\}_{k \times \ell} \middle| p_{ij} \geq 0 \text{ for all } i, j; \sum_{i,j} p_{ij} = 1 \right\} \)

Coefficient \( A \) is a map

\[ A : M \rightarrow \mathbb{R} \]

(codomain is usually \([0, 1]\) or \([-1, 1]\))

Coefficient space

\[ D = \{ A : M \rightarrow \mathbb{R} \} \]

will be used as the domain of

- correction for chance function
- correction for maximum value function
Correction for chance function

Function $c : D \to D$, 

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

where

- $A$ is coefficient
- $E(A)$ is value under chance (conditionally upon fixed marginal totals)
- $M(A)$ is overall maximum value

Example: if $A = O = \sum_{i=1}^{k} p_{ii}$

then $c(A) = c(O) = \frac{\sum_{i=1}^{k} (p_{ii} - p_{i+}p_{+i})}{1 - \sum_{i=1}^{k} p_{i+}p_{+i}} = \kappa$
Lemma: Let $A, B \in D$ such that $B = a + bA$, with $a, b \in \mathbb{R}$ and $b \neq 0$. Then $c(A) = c(B)$.

Proof: $E(B) = a + bE(A)$ and $M(B) = a + bM(A)$ and thus

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

Lemma: Let $A, B \in D$ such that $c(A) = c(B)$. Then $c(A + B) = c(A) = c(B)$. 

Matthijs J Warrens  Correction for chance and correction for maximum value
Coefficients that coincide

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

Corollary: two linear transformations of $A$ coincide if they have the same ratio

\[
\frac{M(B) - \lambda}{\mu}
\]

(Albatineh et al. 2006)

Results help identify which coefficients coincide after correction for chance
Correction for maximum value function

Function \( d : D \to D, \quad d(A) = \frac{A}{m(A)} \)

where

- \( A \) is the coefficient
- \( m(A) \) is maximum value given the marginal totals

Example: maximum value of Cohen’s kappa given the marginal totals is

\[
m(\kappa) = \frac{\sum_i (\min \{p_{i+}, p_{+i}\} - p_{i+p}+p_{+i})}{1 - \sum_i p_{i+p}+p_{+i}}
\]

Thus

\[
d(\kappa) = \frac{\kappa}{m(\kappa)} = \frac{\sum_i (p_{ii} - p_{i+p}+p_{+i})}{\sum_i (\min \{p_{i+}, p_{+i}\} - p_{i+p}+p_{+i})} = H
\]
Composition

Two compositions

\[ cd(A) = c(d(A)) \quad \text{and} \quad dc(A) = d(c(A)) \]

Lemma: \( cd = dc \)

Final result does not depend on the order in which functions are applied

Correction for chance function and correction for maximum value function commute
Linear transformations

Lemma: Let $A, B \in D$ such that $B = \lambda + \mu A$, where $\lambda$ and $\mu \neq 0$ are functions of the marginal totals. Then $cd(B) = cd(A)$.

All linear transformations of $A$ coincide after two corrections.

There is precisely one linear transformation of $A$ that has

- value 0 under statistical independence
- maximum value 1 regardless of marginal distributions

Matthijs J Warrens
Lemma: \( c^2 = c \)

_Proof:_

\[
E(c(A)) = \frac{E(A) - E(A)}{M(A) - E(A)} = 0,
\text{ thus } c(c(A)) = \frac{c(A) - 0}{1 - 0} = c(A)
\]

Lemma: \( d^2 = d \).

Lemma: \( (cd)^2 = cd \).

_Proof:_ \( cdcd = c^2d^2 = cd \)
Let \( 1 : D \to D \) denote the identity function.

Combining the lemmas yields the table:

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>c</th>
<th>d</th>
<th>cd</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>c</td>
<td>d</td>
<td>cd</td>
</tr>
<tr>
<td>c</td>
<td>c</td>
<td>c</td>
<td>cd</td>
<td>cd</td>
</tr>
<tr>
<td>d</td>
<td>d</td>
<td>cd</td>
<td>d</td>
<td>cd</td>
</tr>
<tr>
<td>cd</td>
<td>cd</td>
<td>cd</td>
<td>cd</td>
<td>cd</td>
</tr>
</tbody>
</table>

Function \( cd \) acts as an absorbing element.

Let \( R = \mathbb{Z} \setminus 2\mathbb{Z} \) be the ring of integers modulo 2. 
\( \{1, c, d, cd\} \) is isomorphic to \( R^2 \) (multiplication componentwise).
The set \{1, c, d, cd\} is

- closed under multiplication (function composition)
- associative
- has an identity element

Hence, it is an idempotent commutative monoid