

Stochastic orderings and their use in statistics: the case of association between two variables

Ordinamenti stocastici e loro uso in statistica: il caso della associazione tra due variabili

Alessandra Giovagnoli ⁽¹⁾

Dipartimento di Scienze Statistiche “Paolo Fortunati”

Università di Bologna

giovagno@stat.unibo.it

Riassunto: Il concetto di associazione tra variabili categoriche nominali (connessione) viene qui identificato da una relazione di pre-ordinamento. Sono possibili diversi ordinamenti che esprimono connessione sia quando le variabili intervengono simmetricamente (caso dell’interdipendenza), che nel caso di dipendenza di una delle due dall’altra. Si cerca di rispondere alle domande: 1. Quand’è che, rispetto a un particolare concetto di associazione, una coppia di variabili è “ugualmente associata” oppure “meno associata” di un’altra coppia di variabili? Quali operazioni sulle variabili lasciano inalterata quella particolare relazione di associazione tra esse? Quali la aumentano o la diminuiscono? 2. Quali sono gli indicatori che rispettano l’ordinamento relativo a quel particolare legame associativo? La presentazione si focalizza poi sul *raters’ agreement*. Il “grado di accordo” tra due osservatori può essere visto come un particolare tipo di connessione tra le variabili che rappresentano le osservazioni di due diversi esperti che in parte concordano e in parte no. L’esempio che mi motiva è quello degli esami citologici di laboratorio.

Keywords: Categorical Variables, Cohen’s kappa, Cross-classification Tables, Dependence, Stochastic Matrices.

1. Introduction: defining association

Interest in assessing the association between two statistical or random variables, meaning the strength of their mutual dependence or dependence of one on the other, has a long history in the statistical literature and a fundamental survey is that by Goodman and Kruskal (1979). Usually the interest lies in how to *measure* the association: several measures are possible, which reflect different goals we may have in mind for specific applications (Haberman, 1982). This paper, on the other hand, focuses on describing and comparing various *concepts* of association, a mental operation which logically seems to come before resorting to a given measure. Many concepts in statistics (association is one of them, but the same could be said of variability, concentration . . .) are not amenable to a direct definition, much as we are at a loss when we try to explain the idea of beauty, but just as for beauty, we can define them through a series of comparisons. In fact, a pair of variables (X, Y) is not described just by possessing the attribute in question, i.e. being or not as-

¹The author wishes to thank the Cytology Chair of the Second School of Medicine, “La Sapienza” University, Rome, for kindly providing the clinical references and the data. She also wishes to thank Professor Antonio Forcina for useful comments and Dr Simone Giannerini for technical help. The research is sponsored by the CNR Project “Metodi e Sistemi per il Supporto alle Decisioni” (MSSD).

sociated, but also by possessing it to a greater or lesser extent. The concept is identified by an order relation (usually a partial ordering, since not all bivariate distributions can be compared) and its strength will be measured by indicators that are thus order-preserving functions. However, the introduction of indicators makes all comparisons possible, and one may say that this will sometimes be unnatural. An axiomatic investigation of statistical concepts by means of order and equivalence relations is suggested by Bickel and Lehmann (1975), while an elegant presentation for teaching purposes has been given by Regazzini (1987).

Speaking quite in general, there are different types of association. A first fundamental distinction is between *interdependence* (the two variables have a symmetrical role) and *dependence* of one on the other. Another important distinction is between association with or without a “direction”: in both cases a cross-classification table which refers to independence stands for the null association, but the independence table may be a minimum w.r.t. the ordering or simply an intermediate case between tables with positive and tables with negative association. I shall mark the distinction by coining the phrase *directional association* to describe the latter situation. Furthermore, the 2×2 case often presents special features and may deserve to be looked at separately. The literature on association orderings is extensive but it mainly deals with ordered and quantitative variables whereas in this paper I intend to examine order relations for association of **nominal** categorical variables. Far from attempting an exhaustive review of all the possible concepts of association in this context, I will mention a few and for them try to partially answer the following questions:

1. a) When is a pair of variables as associated as another pair with respect to a given concept of association? In other words, how can we describe the equivalence relation “the two pairs of variables are equally associated” with reference to that particular aspect? b) Which operations on a table P , expressing the joint distribution of the two variables, do not affect that particular type of association among them?
2. a) When is a pair of variables more associated than another pair with respect to a given concept of association? In other words, how can we describe the order relation “one pair of variables is more associated than the other” with reference to that particular aspect? b) Which operations on a pair of variables increase or decrease that particular type of association among them?
3. What are suitable criteria for representing the suggested ordering on a numerical scale, i.e. what are the coherent indicators for that type of association?

My presentation will then concentrate on a special example of directional association, that of *agreement among raters*. In this case there is only one characteristic of interest, observed on the same statistical units by two different observers who may partly agree and partly disagree, and the two variables represent the observations of the two “raters”. As far as I know there are no known results for when the strength of the agreement between the two observers is assessed not by a single measurement, as is customary in the literature, but by an order relation.

For reasons of space the ideas put forward in this paper are just sketched out and the proofs are only hinted at.

2. Some terminology

Let us recall some definitions. An *equivalence* in a set $\mathcal{S}$ is a reflexive, symmetric, transitive relation. A *pre-order* $\leq$ in $\mathcal{S}$ is a reflexive and transitive relation (anti-symmetry is not required). To every pre-order $\leq$ there corresponds an equivalence relation $\triangleq$: if $x, y \in \mathcal{S}$ and $x \leq y, y \leq x$ then we say that $x \triangleq y$. All the one-to-one maps φ of $\mathcal{S}$ onto $\mathcal{S}$ such that $\varphi(x) \triangleq x$ form a set $\mathbf{G}_I$, called the *invariance* set of $\leq$. All the maps ψ of $\mathcal{S}$ into $\mathcal{S}$ such that $x \leq y$ implies $\psi(x) \leq \psi(y)$ form the *equivariance* set $\mathbf{G}_E$ of $\leq$. All the maps ϕ of $\mathcal{S}$ into $\mathcal{S}$ such that $\phi(x) \leq x$ for all $x \in \mathcal{S}$ form the *contraction* set $\mathbf{G}_K$ of $\leq$. As well as the contractions, we can define the *expansions*: all the maps $\tilde{\phi}$ of $\mathcal{S}$ into $\mathcal{S}$ such that $x \leq \tilde{\phi}(x)$ for all $x \in \mathcal{S}$. Clearly $\mathbf{G}_I \subset \mathbf{G}_E$ and $\mathbf{G}_I \subset \mathbf{G}_K$; $\mathbf{G}_I$ is a group, whereas $\mathbf{G}_E$ and $\mathbf{G}_K$ need not be.

A function $f : \mathcal{S} \rightarrow \mathbb{R}$ is *order-preserving* if $x \leq y$ implies $f(x) \leq f(y)$. A trivial remark: if f is order-preserving and $g : \mathbb{R} \rightarrow \mathbb{R}$ is non-decreasing, then $g \circ f$ too is order-preserving. Finding $\mathbf{G}_I$, $\mathbf{G}_E$, and $\mathbf{G}_K$ for a given $\leq$ means answering questions 1b and 2b of the previous Section. Finding the class $\mathfrak{S}$ of all order preserving functions is the full answer to question 3. Clearly order-preserving functions must be invariant w.r.t. $\mathbf{G}_I$.

Let X and Y be categorical variables with a finite number of unordered categories, denoted by $x_1, \dots, x_r$ and $y_1, \dots, y_c$ respectively, and let

$$P_{r \times c} = \begin{pmatrix} p_{11} & p_{12} & \dots & \cdot & p_{1c} \\ p_{21} & p_{22} & \cdot & \cdot & p_{2c} \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ p_{r1} & \cdot & \cdot & \cdot & p_{rc} \end{pmatrix}$$

be their joint distribution table (the p_{ij} 's may be frequencies or probabilities). Thus we want to compare sets of triplets $\{(x_i, y_j, p_{ij}); i = 1, \dots, r; j = 1, \dots, c\}$ (see Zanella, 1988) defined up to a permutation of the i 's and one of the j 's. When the x 's and y 's are fixed, it is sufficient to refer to the matrix P . An alternative description is by means of the conditional and marginal distributions i.e. either $(P^*, \mathbf{p}_r)$, where $P^* = \left(\frac{p_{ij}}{p_{i\cdot}} \right)$ and $\mathbf{p}_r = (p_{1\cdot}, \dots, p_{r\cdot})^T$, or $(P^{**}, \mathbf{p}_c)$, where $P^{**} = \left(\frac{p_{ij}}{p_{\cdot j}} \right)$ and $\mathbf{p}_c = (p_{\cdot 1}, \dots, p_{\cdot c})^T$. It is sometimes useful to include tables with null rows or columns, in which case P^* or P^{**} are not uniquely defined, but no serious inconvenience is experienced as a result.

3. Dependence orderings for two categorical variables

I start by briefly recalling some results obtained in Forcina and Giovagnoli (1987) on orderings which refer to the dependence of a variable Y on another one X , both categorical. The order relation $\leq_S$ is defined by

$$Q \leq_S P \stackrel{def}{\iff} Q = S^t P \quad (1)$$

with S a stochastic matrix. It implies that the column margins are equal: $\mathbf{p}_c = \mathbf{q}_c$, and it can be seen to be equivalent to the following two conditions:

$$\text{i) } Q^* = \tilde{S} P^*; \quad \text{ii) } \mathbf{q}_r = \tilde{S}^t \mathbf{p}_r$$

with $\tilde{S}$ another stochastic matrix. The equivalence relation $\simeq_S$ defined by $\leq_S$ is permutation of the rows of P . In Forcina and Giovagnoli (1987) we showed that $\leq_S$ satisfies some intuitive requirement for a dependence ordering. For instance it can be proved that:

A1 Row-aggregation, i.e. replacing one of two rows by their sum and the other by a row of zeros, decreases $\leq_S$. If the rows are proportional, row-aggregation and row splitting lead to equivalence under $\simeq_S$.

This follows because $\left(\begin{array}{cc|c} 1 & 1 & 0 \\ 0 & 0 & \\ \hline 0^t & & I_{r-2} \end{array} \right)$ and $\left(\begin{array}{cc|c} \alpha & 1 & 0 \\ 1-\alpha & 0 & \\ \hline 0^t & & I_{r-2} \end{array} \right)$ are transposed stochastic matrices.

Equally, if we were interested in comparing the dependence of X on Y we could consider the “transpose” of $\leq_S$, namely

$$Q \leq_S^t P \stackrel{def}{\iff} Q = PS, S \text{ stochastic.} \quad (2)$$

Another dependence order is

$$Q \leq_D P \stackrel{def}{\iff} Q = PD \quad (3)$$

with D a product of T -transfers, namely matrices of the form $T_\alpha = (1 - \alpha)I + \alpha\Pi^{(2)}$ with $0 \leq \alpha \leq 1$ and $\Pi^{(2)}$ a permutation matrix that exchanges only 2 elements; D is doubly-stochastic. This can be thought of as a model for errors in the Y variable: α is the probability (frequency) of mistaking two y -categories. This ordering is known in the literature as chain-majorization (see Marshall and Olkin, 1979). The equivalence relation $\simeq_D$ defined by $\leq_D$ is permutation of the columns of P . The invariance and contraction sets of $\leq_S$ and $\leq_D$ are as follows, where the pair (A, B) denotes two matrices of dimensions $r \times r$ and $c \times c$ respectively, acting on P respectively by pre- and by post-multiplication.

1. $\mathbf{G}_I(\leq_S) = \{(\Pi_1, I_c); I_c \text{ the identity, } \Pi_1 \text{ a permutation matrix}\}$
2. $\mathbf{G}_I(\leq_D) = \{(I_r, \Pi_2); I_r \text{ the identity, } \Pi_2 \text{ a permutation matrix}\}$
3. $\mathbf{G}_K(\leq_S) = \{(S^T, I_c); I_c \text{ the identity, } S \text{ stochastic}\}$
4. $\mathbf{G}_K(\leq_D) = \{(I_r, D); I_r \text{ the identity, } D \text{ a } T\text{-matrix}\}$

Furthermore it can be shown that

5. $\mathbf{G}_E(\leq_S) \supseteq \{(S_1^T, S_2); S_1 \text{ stochastic and of full rank, } S_2 \text{ stochastic}\}$
6. $\mathbf{G}_E(\leq_D) \supseteq \{(S^T, D); S \text{ stochastic, } D \text{ a } T\text{-matrix}\}$

Clearly these two orderings can be combined to define the order relation that we can call *FG-dependence*

$$Q \leq_{FGdep} P \stackrel{def}{\iff} Q = S^t PD \quad (4)$$

for some stochastic matrix S and some T -matrix D . This means that there exists a bivariate distribution table R such that $R \leq_S P$ and $Q \leq_D R$ (and also there exists $\tilde{R}$ such that $\tilde{R} \leq_D P$ and $Q \leq_S \tilde{R}$). It can be written as $Q = (S \otimes D)^t \text{vec}(P)$.

As to the indicators of dependence used in the literature, in general they are of two types: those based on comparing optimal prediction of Y given X with optimal prediction of Y when X is unknown, and the ones that measure the mean information of Y given X relative to unconditional information of Y . We write, for the first and second type respectively,

$$\text{I)} \quad \Phi_{Y.X} = \frac{\Delta(\mathbf{p}_c) - \sum_i p_i \Delta(\mathbf{p}_i^*)}{\Delta(\mathbf{p}_c)}$$

where $\Delta(\cdot)$ stands for a measure of dispersion/heterogeneity of the distribution of Y , or of minimal expected loss in predicting Y .

$$\text{II)} \quad \Psi_{Y.X} = \sum_i p_i g[d(\mathbf{p}_i^*, \mathbf{p}_c)]$$

where $g : \mathbb{R} \rightarrow \mathbb{R}$ is one-one monotonic and $d(\mathbf{u}, \mathbf{v})$ is a measure of the “distance” of distribution $\mathbf{u}$ from $\mathbf{v}$, or of information gain from prior $\mathbf{v}$ to posterior $\mathbf{u}$.

All measures of FG -dependence must be invariant with respect to permutations of rows and of columns of cross-classification tables. The orders $\leq_S$ and $\leq_D$ are preserved by Type I indices when Δ is convex and permutation invariant. Examples are Guttman’s λ , Goodman and Kruskal’s τ and Theil’s index η . The orders $\leq_S$ and $\leq_D$ are preserved by Type II indices when g is convex, $d(\mathbf{u}, \mathbf{v})$ is invariant under the same T-transform of $\mathbf{u}$ and $\mathbf{v}$ and convex in the first component. Examples are Gini’s connection index $\frac{1}{2} \sum_i \sum_j |p_{ij} - p_i \cdot p_{\cdot j}|$, Good’s class of measures J_λ (which include Pearson’s familiar χ^2) and Halphen’s *modulus of dependence* $\sum_i \sum_j p_{ij} \log p_{ij} - \sum_i p_i \cdot \log p_i - \sum_j p_{\cdot j} \log p_{\cdot j}$.

4. Orderings of non-directional association

Association between two variables in the “symmetric” case may be defined as some type of “distance” from the reference situation of independence, which corresponds to no association. A natural requirement is again invariance with respect to permutations of the rows and of the columns. One possible way of introducing association orders between bivariate distributions which makes use of the heuristic arguments presented in Forcina and Giovagnoli (1987) for the dependence case is combining the dependence ordering $\leq_S$ with its transpose or combining $\leq_D$ with its transpose. Thus we can define

$$Q \leq_{W-ass} P \stackrel{def}{\iff} Q = S_1^t P S_2 \quad (5)$$

$$i.e. \stackrel{def}{\iff} Q = (S_1 \otimes S_2)^t vec(P) \quad (6)$$

for some stochastic matrices S_1 and S_2 . One may immediately show, for instance, that since the independence matrix with the same margins as P is $P \mathbf{1} \mathbf{1}^t P$ and $\mathbf{1} \mathbf{1}^t P$ is stochastic, then $\mathbf{1} \mathbf{1}^t P \leq_{W-ass} P$ for all P . If both P and Q have uniform margins then S_1 and S_2 must be doubly stochastic. I am not aware of any thorough investigation carried out on this order relation.

Also of particular interest is a suggestion made by Cifarelli and Regazzini (1986) for distributions with the same margins, namely regarding association as the concentration of a bivariate distribution w.r.t. its independence table. Let $\eta_{ij} = \frac{p_{ij}}{p_i \cdot p_{\cdot j}}$ be the likelihood ratio of P and $v(x_i, y_j) = p_i \cdot p_{\cdot j}$ the independence distribution. Consider a new discrete

variable η_P and let $F_P(x) = v \{(x_i, y_j) \text{ s.t. } \eta_{ij} \leq x\}$. Observe that $E_v(\eta_P) = 1$. Then we define the association order $\preceq^A$ according to Cifarelli and Regazzini

$$Q \preceq^A P \stackrel{\text{def}}{\iff} \eta_P \leq_L \eta_Q \quad (7)$$

where $\leq_L$ is the concentration (Lorenz) order for discrete variables (see for instance Marshall and Olkin, 1979). Scarsini (1991) proves the following

- The order $\preceq^A$ is invariant under any permutation of the rows and any permutation of the columns.
- Row aggregation decreases $\preceq^A$ and leads to equivalence if the rows are proportional.
- The independence table is the unique minimum w.r.t. $\preceq^A$. Diagonal tables are maxima.

Cifarelli and Regazzini (1986) also obtain a class of indices for $\preceq^A$ of the form

$$\sum_i \sum_j p_{i \cdot} p_{\cdot j} g \left(\frac{p_{ij}}{p_{i \cdot} p_{\cdot j}} \right),$$

with $g(\cdot)$ convex on $[0, \infty)$, which includes Pearson's chi-squared when $g(x) = x^2 - 1$.

If the marginals are uniform, the independence distribution is uniform too on the set $\{(x_i, y_j), i = 1, \dots, r; j = 1, \dots, c\}$. It is well-known that when v is uniform the Lorenz order is majorization (Marshall and Olkin, 1979), thus in this case $\preceq^A$ becomes majorization $\preceq$ of the *vecs*, namely the association order defined by Joe (1985). Hence:

$$Q \preceq^A P \iff \text{vec}(Q) = \mathbf{D} \text{vec}(P)$$

with $\mathbf{D}$ an $rc \times rc$ doubly stochastic matrix. A special case is $\mathbf{D} = (D_1 \otimes D_2)^t$, thus for matrices with uniform marginals $Q \leq_{W-ass} P$ implies $Q \preceq^A P$ and $Q \preceq^A P \iff \text{vec}(Q) \preceq \text{vec}(P)$.

5. A special case of directional association: raters' agreement

Association of a directional type is almost always used in the literature with reference to ordered variables (X, Y) . In such a context one defines *concordance* (positive or negative) orderings and various type of positive quadrant dependence, to be found in (Joe, 1997, chap. 2, section 2). Also of particular interest is the case when one variable indicates time, on a discrete scale, and the other is stochastically increasing (or decreasing) with time. Two different examples in the context of bio-statistical applications have been considered in Baldi Antognini and Giovagnoli (2002) and in Bonetto (2001). The ordering relative to such temporal dependence is called *monotone regression dependence* (denoted $\preceq_{SI}$ in Joe 1997. When the variables are both quantitative, directional association is often defined in terms of covariance or correlation of suitable functions of X and Y (Shaked and Shanthikumar, 1994, chap. 9).

In the case of interrater agreement that I am going to consider in detail now, the categories are NOT necessarily ordered. Suppose two observers independently categorize items or responses among the same set of nominal categories, and we wish to compare the degree of agreement among pairs of observers (Bishop *et al.*, 1975, chap. 11). In

this case all the contingency (or joint distribution) tables will be square ones. Applications range from beauty contests to food tasting, to medical diagnoses, etc. When one of the observers, say the first one, is the same each time and is taken as some sort of reference (the “objective truth”), the problem may be described as one of comparing the reliability of different assessors and can be thought of as the dependence of the variable “assessment” on the “truth”.

Given the table

$$P_{h \times h} = \begin{pmatrix} p_{11} & p_{12} & \dots & \cdot & p_{1h} \\ p_{21} & p_{22} & \cdot & \cdot & p_{2h} \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ p_{h1} & \cdot & \cdot & \cdot & p_{hh} \end{pmatrix},$$

since the categories for the rows will have to appear in the same order as those for the columns, the invariance group will be $\mathbf{G}_I = \{(\Pi, \Pi^t); \Pi \text{ an } r \times r \text{ permutation matrix}\}$. A naive ordering may be the one that compares the elements p_{ii} on the main diagonal, namely the frequencies (probabilities) of agreement of the two observers about category i . Let us define

$$Q \leq_{NAIF} P \iff q_{11} \leq p_{11}, q_{22} \leq p_{22}, \dots, q_{hh} \leq p_{hh}.$$

Ways of measuring the degree of raters’ agreement are reviewed in Banerjee *et al.* (1999). Earlier approaches focused on the total proportion of agreement: $Tpa = \sum_i p_{ii}$. This measurement does not account for the fact that a certain amount of agreement may be expected by chance alone. Cohen’s *kappa* has been proposed as a chance-corrected measure, to discount the observed proportion of agreement by its expected level given the observed marginal distributions:

$$\kappa = \frac{\sum_i p_{ii} - \sum_i p_{i\cdot} p_{\cdot i}}{1 - \sum_i p_{i\cdot} p_{\cdot i}}.$$

Clearly $\kappa = 1 - \frac{1 - \sum_i p_{ii}}{1 - \sum_i p_{i\cdot} p_{\cdot i}}$. This is by far the most widely used indicator, although surrounded by much controversy, and a weighted version exists when the categories are ordered.

We point out that the $\leq_{NAIF}$ ordering is preserved by Tpa but not necessarily by κ , as the following counter-example shows

$$Q = \begin{pmatrix} 0.1 & 0.1 \\ 0.7 & 0.1 \end{pmatrix}; P = \begin{pmatrix} 0.2 & 0.3 \\ 0.4 & 0.1 \end{pmatrix}.$$

6. An application to diagnostic tests

My theoretical framework will be that of cytological tests, e.g. the Papanicolau test for detecting cervical neoplastic lesions. The experimental design to measure the concordance of two screeners A and B demands that each smear be classified independently by the observers according to the same protocol on a scale of h categories; in the case of the Pap test, these may be: Inadequate, Negative, Low Grade (LG), High Grade (HG). The two observers are not necessarily people but may be diagnostic instruments, cytological

or hystological examinations. In some cases “observer” A stands for the so-called Gold Standard (the “true” diagnosis, or the modal one) whereas in other cases there is no Gold Standard. Of particular relevance for clinical implications is the special case of just two categories e.g. Positive or Negative used in primary screening. We now look at the indices usually employed for evaluating interrater agreement (see Fleiss, 1981).

Case I: With the Gold Standard:

The amount of agreement of diagnosis B with A indicates the accuracy of the test. It makes sense to compare only tables with the same row margin. Let P and Q be two $h \times h$ tables with $p_{1\cdot} = q_{1\cdot}, p_{2\cdot} = q_{2\cdot}, \dots, p_{h\cdot} = q_{h\cdot}$, i.e. $P\mathbf{1} = Q\mathbf{1}$. When $h = 2$, let us set: 1 = the test is positive and 2 = the test is negative. Then $P = \begin{pmatrix} p_{11} & p_{12} \\ p_{21} & p_{22} \end{pmatrix}$ with

$$\begin{aligned} p_{11} &= \Pr(\text{True-Positives}) & ; & \quad p_{12} = \Pr(\text{False-Negatives}) \\ p_{22} &= \Pr(\text{True-Negatives}) & ; & \quad p_{21} = \Pr(\text{False-Positives}). \end{aligned}$$

Indicators of accuracy in this case are:

- a) *sensitivity*, defined as $\frac{\Pr(\text{True-Positives})}{\Pr(\text{True-Positives}) + \Pr(\text{False-Negatives})} = \frac{p_{11}}{p_{1\cdot}}$,
and the related percentage of false-negatives $\frac{\Pr(\text{False-Negatives})}{\Pr(\text{True-Positives}) + \Pr(\text{False-Negatives})}$.

Clearly one must not forget the different relevance of these parameters to specific pathologies.

- b) *specificity* = $\frac{\Pr(\text{True-Negatives})}{\Pr(\text{True-Negatives}) + \Pr(\text{False-Positives})} = \frac{p_{22}}{p_{2\cdot}}$.
c) *positive predictive value* = $\frac{\Pr(\text{True-Positives})}{\Pr(\text{True-Positives}) + \Pr(\text{False-Positives})} = \frac{p_{11}}{p_{1\cdot}}$.

Result 1. In 2×2 tables with $P\mathbf{1} = Q\mathbf{1}$, $\text{sensitivity}(Q) \leq \text{sensitivity}(P)$ and $\text{specificity}(Q) \leq \text{specificity}(P)$ if and only if $Q \leq_{NAIF} P$. Moreover $Q \leq_{NAIF} P$ implies positive predictive value of $Q \leq$ positive predictive value of P .

Proof. Elementary. □

Case II: Without Gold Standard

This corresponds to comparing a type of interdependence among observers A and B. Apart from Cohen’s kappa, when $h = 2$ the following agreement measures are used:

- d) *Specific concordance on positives*: this is the frequency (or probability) of the cases that having been diagnosed as positive by at least one of the two observers are also positive according to the other, and is calculated by

$$C^+ = \frac{2p_{11}}{2p_{11} + p_{21} + p_{12}}$$

- e) *Specific concordance on negatives*:

$$C^- = \frac{2p_{22}}{2p_{22} + p_{21} + p_{12}}.$$

However, when corrected for chance as explained in section 5, both these indicators coincide with Cohen’s kappa.

Result 2.

$$P \leq_{NAIF} Q \implies C^+(P) \leq C^+(Q) \quad \text{and} \quad C^-(P) \leq C^-(Q).$$

Proof. It is easy to see that from the point of view of comparing P and Q , $C^+(P)$ and $C^-(P)$ are equivalent to considering $p_{11}/(1-p_{22})$ and $p_{22}/(1-p_{11})$, hence the proof. $\square$

The converse does not hold as the following tables show:

$$P = \begin{pmatrix} 0.4 & 0.1 \\ 0.1 & 0.4 \end{pmatrix}, \quad Q = \begin{pmatrix} 0.30 & 0.05 \\ 0.05 & 0.60 \end{pmatrix}.$$

For reasons of space this preliminary investigation of agreement orderings has to stop at this seminal stage. There seems to be a need to go beyond the $\leq_{NAIF}$ relation, especially when $h > 2$, and find more interesting order relations that express raters' agreement as a particular case of row-column association.

References

- Baldi Antognini A. and Giovagnoli A. (2002) A new 'biased coin design' for the sequential allocation of two treatments, manuscript.
- Banerjee M., Capozzoli M. and McSweeney L. (1999) Beyond kappa: A review of interrater agreement measures, *The Canadian Journal of Statistics*, 27, 3–23.
- Bickel P. and Lehmann E.L. (1975) Descriptive statistics for non-parametric models. 1: Introduction, *The Annals of Statistics*, 3, 1038–1044.
- Bishop Y.M.M., Fienberg S. and Holland P. (1975) *Discrete multivariate analyses: Theory and practice*, MIT Press, Cambridge, MA.
- Bonetto C. (2001) *Metodi geometrici per l'analisi di tavole di contingenza con dati ordinali: Un'applicazione in ambito ospedaliero*, Ph.D. dissertation, Dipartimento di Scienze Statistiche, Università di Bologna.
- Cifarelli D.M. and Regazzini E. (1986) Concentration function and its role in descriptive statistics, in: *Proceedings of the 33rd scientific meeting of the Italian statistical society*, volume 2, 347–352, in Italian.
- Fleiss J.L. (1981) *Statistical methods for rates and proportions*, John Wiley & sons, New York.
- Forcina A. and Giovagnoli A. (1987) Mathematical properties of cross-classification matrices, *Boll. U.M.I.*, VII, 1-B, 2, 523–544.
- Goodman L.A. and Kruskal W. (1979) *Measures of association for cross classifications*, Springer-Verlag Inc, Berlin, New York.
- Haberman S.J. (1982) Measures of association, in: *Encyclopedia of statistical sciences*, John Wiley & sons, volume 1, chapter 1, 130–137.
- Joe H. (1985) An ordering of dependence for contingency tables, *Linear Algebra and its Applications*, 70, 89–103.
- Joe H. (1997) *Multivariate models and dependence concepts*, Chapman & Hall Ltd, London, New York.
- Marshall A.W. and Olkin I. (1979) *Inequalities – Theory of majorization and its applications*, Academic Press, London, New York.
- Regazzini E. (1987) *Argomenti di Statistica, 2. Variabilità - Concentrazione - Connessione - Concordanza - Dipendenza*, Dipartimento di Matematica, Università di Milano.

- Scarsini M. (1991) An ordering of dependence, in: *Topics in statistical dependence*, Block H., Sampson A.R. and Savits T., eds., Institute of Mathematical Statistics, Hayward, 403–414.
- Shaked M. and Shanthikumar J.G. (Eds.) (1994) *Stochastic orders and their applications*, Academic Press, London, New York.
- Zanella A. (1988) Comparison of association measures with reference to unordered categorized variables, in: *Proceedings of the 34rd scientific meeting of the Italian statistical society*, volume 2, 117–124, in Italian.