The Measurement of Association of Rows and Columns for an $r \times s$ Contingency Table

By PATRICIA M. E. ALTHAM

Medical Research Council and Newnham College, Cambridge

[Received September 1968. Revised February 1969]

SUMMARY

The generalization of Edwards's argument for the measure of association of the rows and columns of 2×2 table, to that of an $r \times s$ table whose rows and columns are assumed unordered, shows, not surprisingly, that association ought to be measured by some function of the $(r-1)(s-1)$ cross-ratios. Such a function is suggested by the introduction of a metric on certain equivalence classes. The properties of such metrics are examined, and in particular comparisons are made with Good's suggestion of the use of the algebraic rank of the contingency table, and with Lindley's significance test for association in the $r \times s$ table.

1. GENERALIZATION OF EDWARD'S ARGUMENT FOR THE 2×2 TABLE

EDWARDS (1963) has shown that any measure of association of rows and columns for a 2×2 contingency table should be a function of the cross-ratio $(p_{11}p_{22})/(p_{12}p_{21})$, where the table is represented by $(p_{ij};\ 1 \leqslant i,j \leqslant 2)$. This argument is generalized below to the $r \times s$ contingency table, and some suggestions for measures of association are given.

Let $P_{r,s}$ be the class of $r \times s$ contingency tables, that is,

$$P_{r,s} = \{(p_{ij}) : p_{ij} > 0, 1 \leqslant i \leqslant r, 1 \leqslant j \leqslant s \quad \text{and} \quad \sum\sum p_{ij} = 1\}$$

It is presumed that no orderings of the rows and columns of tables of $P_{r,s}$ are available.

Let $\overset{A}{\sim}$ be an equivalence relation such that for $p, q \in P_{r,s}, p \overset{A}{\sim} q$ means that the "association" of rows and columns of p is the same as the "association" of rows and columns of q. A precise definition of "association" is deliberately omitted at this point, but following Edwards, two propositions about it are made.

Proposition I

If $p, q \in P_{r,s}$ are such that

$$\frac{p_{ij}}{p_{i.}} = \frac{q_{ij}}{q_{i.}}, \quad 1 \leqslant i \leqslant r, \quad 1 \leqslant j \leqslant s-1,$$

then $p \overset{A}{\sim} q$ (employing the usual dot notation for summation over a suffix).

Proposition II

If $p, q \in P_{r,s}$ are such that

$$\frac{p_{ij}}{p_{.j}} = \frac{q_{ij}}{q_{.j}}, \quad 1 \leqslant i \leqslant r-1, \quad 1 \leqslant j \leqslant s,$$

then $p \overset{A}{\sim} q$.

Let $\overset{CR}{\sim}$ be a relation such that for $p, q \in P_{r,s}$, $p \overset{CR}{\sim} q$ iff

$$\frac{p_{ij}p_{rs}}{p_{is}p_{rj}} = \frac{q_{ij}q_{rs}}{q_{is}q_{rj}}, \quad 1 \leqslant i \leqslant r-1, \quad 1 \leqslant j \leqslant s-1.$$

This is easily seen to be an equivalence relation. The generalization of Edwards's result is the following.

Theorem. If $p, q \in P_{r,s}$ and $p \overset{CR}{\sim} q$, then $p \overset{A}{\sim} q$.

Proof. Define

$$\alpha_{ij} = \frac{p_{ij}p_{rs}}{p_{is}p_{rj}} = \frac{q_{ij}q_{rs}}{q_{is}q_{rj}}, \quad 1 \leqslant i \leqslant r, \quad 1 \leqslant j \leqslant s.$$

Define $\pi_{ij} = \pi_0 \alpha_{ij} p_{is} q_{rj}$, where π_0 is the normalizing constant, i.e. such that $\pi \in P_{r,s}$. Then $\pi_{ij}/\pi_{i.} = q_{ij}/q_{i.}$, $1 \leqslant i \leqslant r$, $1 \leqslant j \leqslant s-1$, so that $\pi \overset{A}{\sim} q$ and $\pi_{ij}/\pi_{.j} = p_{ij}/p_{.j}$, $1 \leqslant i \leqslant r-1$, $1 \leqslant j \leqslant s$, so that $\pi \overset{A}{\sim} p$.

Hence, by the transitivity of $\overset{A}{\sim}$, $p \overset{A}{\sim} q$ as required.

Similar versions of this result for $r \times s \times t$ tables are easily obtained, given suitable modification of Propositions I and II. There are essentially four different types of dependence of rows, columns and layers to be considered, and accordingly four different types of cross-ratio arise as being relevant for the description of these four types of association. An example follows.

Let

$$P_{r,s,t} = \{(p_{ijk}) : p_{ijk} > 0,\ 1 \leqslant i \leqslant r,\ 1 \leqslant j \leqslant s,\ 1 \leqslant k \leqslant t, \text{ and } \sum\sum\sum p_{ijk} = 1\}.$$

Let $\overset{B}{\sim}$ be an equivalence relation such that for p, $q \in P_{r,s,t}$, $p \overset{B}{\sim} q$ means that the second order interaction of the rows, columns and layers of p is the same as that of q. Then three propositions about $\overset{B}{\sim}$ are made.

Proposition I′

If $p, q \in P_{r,s,t}$ satisfy

$$\frac{p_{ijk}}{p_{.jk}} = \frac{q_{ijk}}{q_{.jk}}, \quad 1 \leqslant i < r, \quad 1 \leqslant j \leqslant s, \quad 1 \leqslant k \leqslant t,$$

then $p \overset{B}{\sim} q$.

Proposition II′

If $p, q \in P_{r,s,t}$ satisfy

$$\frac{p_{ijk}}{p_{i.k}} = \frac{q_{ijk}}{q_{i.k}}, \quad 1 \leqslant i \leqslant r, \quad 1 \leqslant j < s, \quad 1 \leqslant k \leqslant t,$$

then $p \overset{B}{\sim} q$.

Proposition III′

If $p, q \in P_{r,s,t}$ satisfy

$$\frac{p_{ijk}}{p_{ij.}} = \frac{q_{ijk}}{q_{ij.}}, \quad 1 \leqslant i \leqslant r, \quad 1 \leqslant j \leqslant s, \quad 1 \leqslant k < t,$$

then $p \overset{B}{\sim} q$. Then, if

$$\frac{p_{ijk} p_{ist} p_{rjt} p_{rsk}}{p_{rst} p_{rjk} p_{isk} p_{ijt}} = \frac{q_{ijk} q_{ist} q_{rjt} q_{rsk}}{q_{rst} q_{rjk} q_{isk} q_{ijt}}, \quad 1 \leqslant i < r, \quad 1 \leqslant j < s, \quad 1 \leqslant k < t,$$

$p \overset{B}{\sim} q$.

A similar version of the generalization of Edwards's result is easily proved for "well-behaved" bivariate density functions, given the appropriate versions of Propositions I and II above. However, taking these two postulates as starting points for the discussion of the association of a pair of continuous-valued random variables is possibly not the appropriate approach. Certainly Rényi (1959) and Ali and Silvey (1965) use different approaches here. However, the function corresponding to the "cross-ratio", that is $\{f(x,y) f(x',y')\}/\{f(x,y') f(x',y)\}$ is considered by Lehmann (1966) as an indication of the dependence of a pair of continuous-valued random variables, with density $f(x,y)$.

Existence and uniqueness of a table $p \in P_{r,s}$, given the $(r-1)(s-1)$ cross-ratios $(p_{ij} p_{rs})/(p_{is} p_{rj})$ and the marginal totals $(p_{i.})$, $(p_{.j})$, where $\sum_i p_{i.} = \sum_j p_{.j} = 1$, follow from the results of Sinkhorn (1967). An alternative proof of uniqueness is given below.

Suppose $p, q \in P_{r,s}, p \overset{CR}{\sim} q$ and $p_{i.} = q_{i.}$, $1 \leqslant i \leqslant r$, $p_{.j} = q_{.j}$, $1 \leqslant j \leqslant s$. Suppose $p \neq q$, and take $p_{11} > q_{11}$ without loss of generality, then since $p_{1.} = q_{1.}$, take $p_{12} < q_{12}$ without loss of generality. Also $p_{.1} = q_{.1}$, so take $p_{21} < q_{21}$ without loss of generality. Now $p_{11} p_{22}/p_{12} p_{21} = q_{11} q_{22}/q_{12} q_{21}$, so $p_{22} < q_{22}$.

Now $p_{2.} = q_{2.}$, and $p_{.2} = q_{.2}$ so, without loss of generality, $p_{23} > q_{23}$, $p_{32} > q_{32}$.

Using $p \overset{CR}{\sim} q$ again, it is apparent that $p_{33} > q_{33}$, $p_{31} > q_{31}$, $p_{13} > q_{13}$. Continuing in this way, a contradiction is reached, for if $r \leqslant s$ it follows that $p_{rj} - q_{rj}$ has the same sign for $1 \leqslant j \leqslant s$, or, if $r \geqslant s$, $p_{is} - q_{is}$ has the same sign for $1 \leqslant i \leqslant r$. Hence $p_{ij} = q_{ij}$, $1 \leqslant i \leqslant r$, $1 \leqslant j \leqslant s$.

This type of argument can be used to show that an $r \times s \times t$ table (p_{ijk}) in $P_{r,s,t}$ is uniquely defined by $(p_{.jk})$, $(p_{i.k})$ and $(p_{ij.})$, provided that these are consistent, and the $(r-1)(s-1)(t-1)$ cross-ratios given above, for $t = 2$. Birch (1963) has demonstrated the existence and uniqueness of a table (p_{ijk}) in $P_{r,s,t}$ given no second-order interaction, and consistent marginal totals $(p_{.jk}), (p_{i.k}), (p_{ij.})$.

The conclusion is that, if Propositions I and II are accepted, then the measure of association of the rows and columns of a table $p \in P_{r,s}$ should be a function of the $(r-1)(s-1)$ cross-ratios $(p_{ij} p_{rs})/(p_{is} p_{rj})$. This conclusion was also reached by Plackett (1962) when considering the second-order interaction in the $r \times s \times t$ table.

It is interesting that the $(r-1)(s-1)$ probability cross-ratios are the only parameters relevant for the distribution obtained by conditioning the $(rs-1)$-dimensional multinomial distribution for a sample of size n from the table on the row and column sums $(n_{i.})$ and $(n_{.j})$.

2. METRICS

As Edwards pointed out, most of the well-known measures of association for the 2×2 contingency table cannot be written as functions of the cross-ratios alone, and this criticism naturally extends to the general $r\times s$ table. For example, the presence of $p_{i.}$ and $p_{.j}$ in such expressions as

$$\Sigma\Sigma\{(p_{ij}-p_{i.}p_{.j})^2/p_{i.}p_{.j}\} \quad \text{or} \quad \Sigma\Sigma p_{ij}\log\{p_{ij}/(p_{i.}p_{.j})\}$$

cannot be eliminated. Plackett notes that similar criticisms apply to coefficients proposed by Goodman and Kruskal (1954, 1959). Good (1965) has suggested that the algebraic rank of the matrix (p_{ij}) should be considered when measuring association, and his suggestion gains some support here from the fact that if $p\overset{CR}{\sim}q$ then (p_{ij}) and (q_{ij}) do have the same rank, though of course the converse does not apply.

It would still be convenient to have a single expression for the measure of association of the rows and columns of association of the rows and columns of a table, rather than the whole set of $(r-1)(s-1)$ cross-ratios. The definition of a metric on the equivalence classes of $\overset{CR}{\sim}$ would seem a possible way of finding a single coefficient of the association.

For $p\in P_{r,s}$, let

$$\tilde{p}=\{q\colon q\in P_{r,s}, q\overset{CR}{\sim}p\},$$

the equivalence class of p with respect to the relation $\overset{CR}{\sim}$.

Let $d(\tilde{p},\tilde{q})$ be a positive-valued function of the $2(r-1)(s-1)$ cross-ratios $(p_{ij}p_{rs})/(p_{is}p_{rj})$ and $(q_{ij}q_{rs})/(q_{is}q_{rj})$ such that $d(\tilde{p},\tilde{q})=d(\tilde{q},\tilde{p})$, $d(\tilde{p},\tilde{q})=0$ iff $p\overset{CR}{\sim}q$, and $d(\tilde{p},\tilde{\pi})+d(\tilde{\pi},\tilde{q})\geqslant d(\tilde{p},\tilde{q})$, for p, q and $\pi\in P_{r,s}$, then $d(\tilde{p},\tilde{q})$ is a metric on the equivalence classes of $\overset{CR}{\sim}$. Let

$$\tilde{\imath}=\{p\colon p\in P_{r,s}, p_{ij}p_{rs}=p_{is}p_{rj}, 1\leqslant i\leqslant r-1, 1\leqslant j\leqslant s-1\}.$$

Then $\tilde{\imath}$ is the "independence class", and $d(\tilde{p},\tilde{\imath})$ is a measure of the association of the rows and columns of the table p. Roughly speaking, the usual coefficients of association given above measure the "distance" of (p_{ij}) from the particular table $(p_{i.}p_{.j})$ rather than from a general independence table. They are the particular values of the measure of the divergence of a table (p_{ij}) from the table (q_{ij}) where $(q_{ij})=(p_{i.}p_{.j})$, whereas the quantity $d(\tilde{p},\tilde{q})$ measures the difference between the *association* of the rows and columns of p and the *association* of the rows and columns of q. If $d(\tilde{p},\tilde{\imath})>d(\tilde{q},\tilde{\imath})$, then the rows and columns of p are more closely associated than those of q, and $d(\tilde{p},\tilde{q})\geqslant d(\tilde{p},\tilde{\imath})-d(\tilde{q},\tilde{\imath})$ by the triangle inequality.

Some examples of metrics $d(\tilde{p},\tilde{q})$ are given below.

Examples. Denote by α_{ijlm} the ratio $(p_{ij}p_{lm}q_{im}q_{lj})/(p_{im}p_{lj}q_{ij}q_{lm})$

(i) $$d_{lm}(\tilde{p},\tilde{q})=\{\Sigma\Sigma|\log\alpha_{ijlm}|^{\nu}\}^{1/\nu} \quad \text{for } \nu\geqslant 1,$$

where the summation $\Sigma\Sigma$ extends over $1\leqslant i\leqslant r$, $1\leqslant j\leqslant s$.

This metric depends on choice of a particular row and column. It is probably preferable to define a metric $d(\tilde{p},\tilde{q})$ which is invariant under similar permutations of rows of p and q and under similar permutations of columns of p and q. This would satisfy the analogue of Rényi's condition (F) for $d(\tilde{p},\tilde{\imath})$. Rényi (1959) gives a list of seven properties which should be satisfied by a measure $\delta(\cdot,\cdot)$ of dependence of two random variables X and Y on a given probability space. Property (F) is: If the Borel-measurable functions $f(\cdot)$ and $g(\cdot)$ map the real axis in a one-to-one way into itself, then $\delta\{f(X), g(Y)\} = \delta(X, Y)$. Also, it is preferable to define a metric $d(\tilde{p},\tilde{q})$ so that the measure of association $d(\tilde{p},\tilde{i})$ is symmetrical with respect to the rows and columns of p; this measure would then satisfy the analogue of Rényi's Condition (B), that $\delta(X, Y) = \delta(Y, X)$.

Metrics fulfilling both these requirements are listed below:

(ii) $$d(\tilde{p},\tilde{q}) = \sum\sum d_{lm}(\tilde{p},\tilde{q}).$$

(iii) $$d(\tilde{p},\tilde{q}) = \max d_{lm}(p,\tilde{q}),$$

where the maximum is taken over $1 \leqslant l \leqslant r$, $1 \leqslant m \leqslant s$.

(iv) $$d(\tilde{p},\tilde{q}) = \{\sum\sum |\log \alpha_{ijlm}|^{\nu}\}^{1/\nu} \quad \text{for } \nu \geqslant 1,$$

where the summation extends over $1 \leqslant i,\ l \leqslant r,\ 1 \leqslant j,\ m \leqslant s$.

If $\nu = 2$, this particular metric has a form familiar from analyses of variance. For, writing $\theta_{ij} = \log p_{ij} - \log q_{ij}$ the expression for $\{d(\tilde{p},\tilde{q})\}^2$ can be simplified as follows:

$$\{d(\tilde{p},\tilde{q})\}^2 = 4rs\left\{\sum\sum \theta_{ij}^2 - \sum \frac{\theta_{i.}^2}{s} - \sum \frac{\theta_{.j}^2}{r} + \frac{\theta_{..}^2}{rs}\right\}.$$

Apart from the factor $4rs$, this is the residual sum of squares in an ordinary two-way analysis of variance, with rs observations (θ_{ij}).

(v) The "discrete" metric is

$$d(\tilde{p},\tilde{q}) = 0, \quad \text{if } p \overset{CR}{\sim} q,$$
$$= 1, \quad \text{otherwise.}$$

(vi) $$d(\tilde{p},\tilde{q}) = \left\{\sum\sum \left|\frac{p_{ij}p_{lm}}{p_{im}p_{lj}} - \frac{q_{ij}q_{lm}}{q_{im}q_{lj}}\right|^{\nu}\right\}^{1/\nu} \quad \text{for } \nu \geqslant 1.$$

(vii) $$d(\tilde{p},\tilde{q}) = \sum\sum x_{ijlm}$$

where

$$x_{ijlm} = 0, \quad \text{if } \alpha_{ijlm} = 1,$$
$$= 1, \quad \text{otherwise.}$$

(viii) Define

$$\rho(x,y) = \left|\frac{x-y}{x+y}\right| \quad \text{for } x, y > 0,$$

then $\rho(x,y) = 0$ iff $x = y$, $\rho(x,y) = \rho(y,x) \geqslant 0$ and $\rho(x,y) + \rho(y,z) \geqslant \rho(x,z)$, for $x, y, z > 0$.

Define

$$d(\tilde{p},\tilde{q}) = \max\left\{\rho\left(\frac{p_{ij}p_{lm}}{p_{im}p_{lj}}, \frac{q_{ij}q_{lm}}{q_{im}q_{lj}}\right)\right\},$$

then $0 \leqslant d(\tilde{p},\tilde{q}) < 1$, and for $r = s = 2$,

$$d(\tilde{p},\tilde{\imath}) = |p_{11}p_{22} - p_{12}p_{21}|/(p_{11}p_{22} + p_{12}p_{21}).$$

This is the modulus of Yule's coefficient of association.

(ix) With $\rho(x,y)$ as above, define

$$d(\tilde{p},\tilde{q}) = \max\left[\rho\left\{\left(\frac{p_{ij}p_{lm}}{p_{im}p_{lj}}\right)^{\frac{1}{2}}, \left(\frac{q_{ij}q_{lm}}{q_{im}q_{lj}}\right)^{\frac{1}{2}}\right\}\right],$$

then $0 \leqslant d(\tilde{p},\tilde{q}) < 1$, and for $r = s = 2$,

$$d(\tilde{p},\tilde{\imath}) = |1 - (p_{12}p_{21}/p_{11}p_{22})^{\frac{1}{2}}|/\{1 + (p_{12}p_{21}/p_{11}p_{22})^{\frac{1}{2}}\}.$$

This is the modulus of Yule's coefficient of colligation.

3. Remarks

The measure of association of the rows and columns of p, $d(\tilde{p},\tilde{\imath})$, gives no indication of the "direction" of the dependence. However, the sign of the association is irrelevant if the rows and columns are not assumed to be ordered.

All the examples given above, except for (vi), are functions of the "ratios of the cross-ratios", i.e. of α_{ijlm}. If a metric $d(\tilde{p},\tilde{q})$ is of this form, it is invariant under truncation of observations. For suppose an experimenter is prevented from having observations directly from contingency tables p and q, and observations noted arise as follows: an object falling in the (i,j)th position of the p or q table is actually observed only with probability θ_{ij} where $0 < \theta_{ij} \leqslant 1$. This is equivalent to sampling from tables α, β of $P_{r,s}$, where $\alpha_{ij} \propto \theta_{ij}p_{ij}$ and $\beta_{ij} \propto \theta_{ij}q_{ij}$. However, with this particular form of $d(\tilde{p},\tilde{q})$, we have $d(\tilde{\alpha},\tilde{\beta}) = d(\tilde{p},\tilde{q})$.

With the obvious notation this may be rewritten as $d(\tilde{p},\tilde{q}) = d(\widetilde{\theta p},\widetilde{\theta q})$, and so, putting $\theta_{ij} = p_{ij}^{-1}$ and q_{ij}^{-1} in turn, and using obvious notation

$$d(\tilde{p},\tilde{q}) = d(\tilde{\imath},\tilde{q}/p) = d(\tilde{\imath},\tilde{p}/q),$$

and so, taking $\tilde{p} = \tilde{\imath}$, $d(\tilde{p},\tilde{\imath}) = d(\tilde{\imath},\tilde{\imath}/p)$.

This final result extends to more general metrics, for example to (vi) above. For write $\theta_{ijlm} = p_{ij}p_{lm}/p_{im}p_{lj}$ and $\phi_{ijlm} = q_{ij}q_{lm}/q_{im}q_{lj}$. Then, if $d(\tilde{p},\tilde{q})$ is a metric such that

$$d(\tilde{p},\tilde{q}) = F(\alpha_{ijlm};\ \beta_{ijlm};\ 1 \leqslant i,l \leqslant r, 1 \leqslant j,m \leqslant s),$$

where F satisfies

$$F(\alpha_{ijlm};\ \beta_{ijlm};\ 1 \leqslant i,l \leqslant r, 1 \leqslant j,m \leqslant s) = F(\alpha_{imlj};\ \beta_{imlj};\ 1 \leqslant i,l \leqslant r, 1 \leqslant j,m \leqslant s)$$

then

$$d(\tilde{p},\tilde{q}) = F\left(\frac{p_{im}^{-1}p_{lj}^{-1}}{p_{ij}^{-1}p_{lm}^{-1}};\frac{q_{im}^{-1}q_{lj}^{-1}}{q_{ij}^{-1}q_{lm}^{-1}};\ 1 \leqslant i,l \leqslant r, 1 \leqslant j,m \leqslant s\right)$$

so that

$$d(\tilde{p},\tilde{q}) = d\left(\frac{\tilde{\imath}}{p},\frac{\tilde{\imath}}{q}\right)$$

and therefore

$$d(\tilde{p},\tilde{\imath}) = d\left(\tilde{\imath},\frac{\tilde{\imath}}{p}\right),$$

taking $\tilde{q} = \tilde{\imath}$.

The use of metrics satisfying this equation gives rise to a situation in which two tables p and q are of different algebraic rank but have identical measures of association of rows and columns. This does not accord well with Good's suggestion. For example, if

$$(p_{ij}) = \lambda \begin{pmatrix} 1 & 1 & 1 \\ 0\cdot5 & 2 & 1 \\ 1\cdot5 & 3 & 2 \end{pmatrix}$$

where λ is the normalizing constant, and if $q_{ij} \propto p_{ij}^{-1}$, then $d(\tilde{p}, \tilde{\imath}) = d(\tilde{q}, \tilde{\imath})$ but the ranks of (p_{ij}) and (q_{ij}) are 2 and 3 respectively.

It is also plausible that if $d(\tilde{p}, \tilde{q})$ is a metric satisfying this equation, and $d(\tilde{p}, \tilde{\imath})$ is not constant for $\tilde{p} \neq \tilde{\imath}$, then two tables of the same rank may have different measures of association of rows and columns. This statement is easy to prove for any particular value of r and s; for example, take $r = s = 4$ and suppose the statement false, in other words, suppose that if rank (p) = rank (q) then $d(\tilde{p}, \tilde{\imath}) = d(\tilde{q}, \tilde{\imath})$.

A contradiction is easily reached.

Take

$$(p_{ij}) = \lambda \begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & 2 & 3 & 4 \\ 2 & 3 & 4 & 5 \\ 3 & 5 & 7 & 9 \end{pmatrix}.$$

where λ is the normalizing constant. Then (p_{ij}) and (p_{ij}^{-1}) are of ranks 2 and 4 respectively, and $d(\tilde{p}, \tilde{\imath}) = d(\tilde{\imath}/p, \tilde{\imath})$, so by the hypothesis, $d(\tilde{p}, \tilde{\imath})$ is constant for all tables of rank 2 or 4. Now take

$$(p_{ij}) = \lambda \begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & 2 & 3 & 4 \\ 4 & 3 & 2 & 1 \\ 2 & 3 & 4 & 5 \end{pmatrix},$$

where λ is the normalizing constant. Then (p_{ij}) and (p_{ij}^{-1}) are of ranks 3 and 4 respectively and $d(\tilde{p}, \tilde{\imath}) = d(\tilde{\imath}/p, \tilde{\imath})$, so by the hypothesis, $d(\tilde{p}, \tilde{\imath})$ is constant for all tables of rank 3 or 4. Hence $d(\tilde{p}, \tilde{\imath})$ is constant for all $\tilde{p} \neq \tilde{\imath}$, which gives the required contradiction. In fact, if $d(\tilde{p}, \tilde{\imath})$ is constant for all $\tilde{p} \neq \tilde{\imath}$, and also a function of the ratios of the cross-ratios, then $d(\tilde{p}, \tilde{q})$ is the discrete metric, defined in example (v).

In the above remarks no use has been made of the fact that $d(\tilde{p}, \tilde{q})$ satisfies $d(\tilde{p}, \tilde{q}) = 0$ iff $p \overset{CR}{\sim} q$, and $d(\tilde{p}, \tilde{\pi}) + d(\tilde{\pi}, \tilde{q}) \geqslant d(\tilde{p}, \tilde{q})$. The only general property of $d(\tilde{p}, \tilde{q})$ that has been used is that it is a function of the cross-ratios of p and q.

This section is concluded by some numerical examples of some of the metrics suggested above, for a few contingency tables.

The metrics evaluated are (iv) for $\nu = 1, 2$, in turn, (vi) for $\nu = 1, 2$, in turn, and (viii). These form the five columns of Table 1, in order.

TABLE 1

Numerical examples of metrics

Metric	(iv) ($\nu = 1$)	(iv) ($\nu = 2$)	(vi) ($\nu = 1$)	(vi) ($\nu = 2$)	(viii)
$d(\tilde{p}, \tilde{q})$	516	52·1	3033	493	0·9998
$d(\tilde{p}, \tilde{\iota})$	332	31·9	1675	349	0·9802
$d(\tilde{p}, \tilde{\sigma})$	420	39·4	1795	351	0·9950
$d(\tilde{p}, \tilde{\pi})$	497	47·2	1943	355	0·9978
$d(\tilde{q}, \tilde{\iota})$	332	31·9	1675	349	0·9802
$d(\tilde{q}, \tilde{\sigma})$	287	27·1	1615	341	0·9802
$d(\tilde{q}, \tilde{\pi})$	320	31·7	1667	339	0·9950
$d(\tilde{\sigma}, \tilde{\iota})$	89	11·1	120	17	0·6000
$d(\tilde{\pi}, \tilde{\iota})$	166	16·7	270	36	0·8000
$d(\tilde{\sigma}, \tilde{\pi})$	156	15·0	259	32	0·8000
$d(\tilde{a}, \tilde{\iota})$	316	30·4	1398	283	0·9756
$d(\tilde{b}, \tilde{\iota})$	141	17·6	284	45	0·8000
$d(\tilde{c}, \tilde{\iota})$	133	13·6	193	27	0·8000
$d(\tilde{d}, \tilde{\iota})$	78	9·2	96	13	0·6000
$d(\tilde{e}, \tilde{\iota})$	258	30·5	1837	398	0·9802

Note that in the first column, for example, $d(\tilde{a}, \tilde{\iota}) > d(\tilde{e}, \tilde{\iota})$, whereas in remaining columns $d(\tilde{a}, \tilde{\iota}) < d(\tilde{e}, \tilde{\iota})$.

The contingency tables are, apart from the normalizing multiplier,

$$p = \begin{pmatrix} 10 & 1 & 1 & 1 \\ 1 & 10 & 1 & 1 \\ 1 & 1 & 10 & 1 \\ 1 & 1 & 1 & 10 \end{pmatrix}, \quad q = \begin{pmatrix} 1 & 1 & 1 & 10 \\ 1 & 1 & 10 & 1 \\ 1 & 10 & 1 & 1 \\ 10 & 1 & 1 & 1 \end{pmatrix},$$

$$\sigma = \begin{pmatrix} 1 & 2 & 1 & 2 \\ 2 & 1 & 2 & 1 \\ 1 & 2 & 1 & 2 \\ 2 & 1 & 2 & 1 \end{pmatrix}, \quad \pi = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 4 & 1 & 2 & 3 \\ 3 & 4 & 1 & 2 \\ 2 & 3 & 4 & 1 \end{pmatrix},$$

$$a = \begin{pmatrix} 9 & 1 & 1 & 1 \\ 1 & 9 & 1 & 1 \\ 1 & 1 & 9 & 1 \\ 1 & 1 & 1 & 9 \end{pmatrix}, \quad b = \begin{pmatrix} 1 & 3 & 1 & 3 \\ 3 & 1 & 3 & 1 \\ 1 & 3 & 1 & 3 \\ 3 & 1 & 3 & 1 \end{pmatrix}, \quad c = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 4 & 1 & 2 & 3 \\ 3 & 4 & 1 & 2 \\ 8 & 7 & 6 & 9 \end{pmatrix},$$

$$d = \begin{pmatrix} 1 & 2 & 1 & 2 \\ 2 & 1 & 2 & 1 \\ 3 & 3 & 3 & 3 \\ 2 & 4 & 2 & 4 \end{pmatrix}, \quad e = \begin{pmatrix} 1 & 10 & 1 & 10 \\ 10 & 1 & 10 & 1 \\ 11 & 11 & 11 & 11 \\ 2 & 20 & 2 & 20 \end{pmatrix}.$$

(Thus p, q, π, a have algebraic rank 4, c has rank 3, and the remaining arrays have rank 2.)

4. Contrasts of Log-ratios and Metrics

Lindley (1964) has shown that if (Φ_{ij}) has a (posterior) Dirichlet distribution with parameters (ν_{ij}), and if $\Theta_{ij} = \log \Phi_{ij}$, $1 \leqslant i \leqslant r$, $1 \leqslant j \leqslant s$, then the joint posterior distribution of contrasts of (Θ_{ij}) is asymptotically multivariate normal. Defining $(c_{ij}^{(\nu)})$, $1 \leqslant \nu \leqslant t$ the set of t contrasts, so that $\sum\sum c_{ij}^{(\nu)} = 0$ for $1 \leqslant \nu \leqslant t$, the random vector $(\sum\sum c_{ij}^{(\nu)} \Theta_{ij}, 1 \leqslant \nu \leqslant t)$ has, asymptotically, a multivariate normal distribution with means $\sum\sum c_{ij}^{(\nu)} \log \nu_{ij}$, and covariances $\sum\sum c_{ij}^{(\nu)} c_{ij}^{(\nu')} \nu_{ij}^{-1}$, $1 \leqslant \nu, \nu' \leqslant t$. Bloch and Watson (1967) have given an improvement to this approximation.

From the Bayesian point of view, it is desirable to know the posterior distribution of $d(\tilde{\Phi}, \tilde{\imath})$, where it is assumed that observations have been made on a table with unknown parameters (ϕ_{ij}) and at least one set of marginal totals for the observations was not held fixed. To apply Lindley's result, it is appropriate to consider metrics $d(\tilde{p}, \tilde{q})$ which are simple functions of contrasts of $(\log p_{ij} - \log q_{ij})$. One such metric is

$$\text{(x)} \qquad d(\tilde{p},\tilde{q}) = \left\{\sum\sum \left(\log \frac{p_{ij} p_{rs} q_{is} q_{rj}}{p_{is} p_{rj} q_{ij} q_{rs}}\right)^2\right\}^{\frac{1}{2}}.$$

The general form of metrics whose squares are sums of squares of contrasts of log ratios is given below. Let X denote rs-dimensional real vector space. Let

$$C = \left\{(c_{ij}) \colon (c_{ij}) \in X, \sum_i c_{ij} = 0, \quad \text{for } 1 \leqslant j \leqslant s, \sum_j c_{ij} = 0 \quad \text{for} \quad 1 \leqslant i \leqslant r\right\}.$$

Let

$$R = \{(r_{ij}) \colon (r_{ij}) \in X, r_{ij} + r_{rs} = r_{is} + r_{rj}, 1 \leqslant i \leqslant r-1, 1 \leqslant j \leqslant s-1\}.$$

The C and R are complementary vector subspaces of X, of dimension $(r-1)(s-1)$ and $(r+s-1)$ respectively. Define scalar multiplication of vectors of X by $(\mathbf{a}, \mathbf{b}) = \sum\sum a_{ij} b_{ij}$ for $\mathbf{a}, \mathbf{b} \in X$. Then C and R are orthogonal subspaces of X, with respect to this scalar multiplication. This is written as $C = R^{\perp}$ or $R = C^{\perp}$. The reader is reminded that if Y is any vector subspace of X, then $Y^{\perp}$ is the vector subspace complementary to Y, and if Z is another vector subspace of X, the statement $Y \subset Z$ is equivalent to the statement $Y^{\perp} \supset Z^{\perp}$. Take $\mu \geqslant 1$ and $\mathbf{c}^{(1)}, \ldots, \mathbf{c}^{(\mu)}$ vectors of X, and write $\alpha_{ij} = \log(p_{ij}/q_{ij})$ for brevity. Then $p \overset{CR}{\sim} q$ if and only if $\boldsymbol{\alpha} \in R$. Define $d(\tilde{p},\tilde{q}) = \{\sum_\nu (\mathbf{c}^{(\nu)}, \boldsymbol{\alpha})^2\}^{\frac{1}{2}}$ where $1 \leqslant \nu \leqslant \mu$ in the summation. Then $d(\tilde{p},\tilde{q})$ is a metric if and only if the linear subspace spanned by $\mathbf{c}^{(1)}, \ldots, \mathbf{c}^{(\mu)}$, denoted by $L(\mathbf{c}^{(1)}, \ldots, \mathbf{c}^{(\mu)})$, contains C.

To prove this assertion, observe first that for any $\mathbf{c}^{(1)}, \ldots, \mathbf{c}^{(\mu)}$, $d(\tilde{p},\tilde{q}) = d(\tilde{q},\tilde{p})$ and $d(\tilde{p},\tilde{q}) \geqslant 0$. Let π be any element of $P_{r,s}$, and define $\beta_{ij} = \log q_{ij}/\pi_{ij}$. Then for any $\mathbf{c}^{(1)}, \ldots, \mathbf{c}^{(\mu)}$,

$$\begin{aligned} d(\tilde{p},\tilde{q}) + d(\tilde{q},\tilde{\pi}) &= \{\sum (\mathbf{c}^{(\nu)}, \boldsymbol{\alpha})^2\}^{\frac{1}{2}} + \{\sum (\mathbf{c}^{(\nu)}, \boldsymbol{\beta})^2\}^{\frac{1}{2}} \\ &\geqslant [\sum \{(\mathbf{c}^{(\nu)}, \boldsymbol{\alpha}) + (\mathbf{c}^{(\nu)}, \boldsymbol{\beta})\}^2]^{\frac{1}{2}} \\ &= \{\sum (\mathbf{c}^{(\nu)}, \boldsymbol{\alpha} + \boldsymbol{\beta})^2\}^{\frac{1}{2}} \\ &= d(\tilde{p},\tilde{\pi}). \end{aligned}$$

Hence for any $\mathbf{c}^{(1)}, \ldots, \mathbf{c}^{(\mu)}$, $d(\tilde{p},\tilde{q})$ obeys the triangle inequality.

Suppose $L(\mathbf{c}^{(1)},\dots,\mathbf{c}^{(\mu)})\supset C$. Then $d(\tilde{p},\tilde{q})=0$ implies that $\sum(\mathbf{c}^{(\nu)},\boldsymbol{\alpha})^2=0$, and hence $\boldsymbol{\alpha}\in\{L(\mathbf{c}^{(1)},\dots,\mathbf{c}^{(\mu)})\}^{\perp}\subset C^{\perp}=R$. Consequently, $d(\tilde{p},\tilde{q})=0$ implies $\boldsymbol{\alpha}\in R$, i.e. $p\overset{CR}{\sim}q$, and so $d(\tilde{p},\tilde{q})$ is a metric.

Suppose $d(\tilde{p},\tilde{q})$ is a metric. Then $d(\tilde{p},\tilde{q})=0$ implies that $p\overset{CR}{\sim}q$, i.e.

$$\boldsymbol{\alpha}\in\{L(\mathbf{c}^{(1)},\dots,\mathbf{c}^{(\mu)})\}^{\perp}$$

implies that $\boldsymbol{\alpha}\in R$. Hence $\{L(\mathbf{c}^{(1)},\dots,\mathbf{c}^{(\mu)})\}^{\perp}\subset R$, and therefore $L(\mathbf{c}^{(1)},\dots,\mathbf{c}^{(\mu)})\supset R^{\perp}=C$. This gives the required result.

The *minimum* value of μ for which $d(\tilde{p},\tilde{q})$ is a metric is $\mu=(r-1)(s-1)$, in which case $(\mathbf{c}^{(1)},\dots,\mathbf{c}^{(\mu)})$ is a *basis* of C. The metric (x) given above corresponds to the basis of C defined by

$$\begin{aligned} c_{lm}^{(\nu)} &= 1, \quad \text{for } (l,m)=(i,j),(r,s),\\ &= 1, \quad \text{for } (l,m)=(i,s),(r,j),\\ &= 0, \quad \text{otherwise}\end{aligned}$$

for $\nu=\nu(i,j)$, $1\leqslant i\leqslant r-1$, $1\leqslant j\leqslant s-1$.

In view of the form of the posterior distribution of (Θ_{ij}), it is appropriate to define a second scalar multiplication for vectors of X, by $\langle\mathbf{a},\mathbf{b}\rangle=\sum\sum a_{ij}b_{ij}\nu_{ij}^{-1}$, for $\mathbf{a},\mathbf{b}\in X$. This is a true scalar multiplication iff $\nu_{ij}>0$ for all i,j. Let $(\mathbf{d}^{(\nu)})$ be a basis of C orthonormal with respect to the operation $\langle\cdot,\cdot\rangle$, i.e.

$$\begin{aligned}\langle\mathbf{d}^{(\nu)},\mathbf{d}^{(\nu')}\rangle &= 1, \quad \text{for } \nu=\nu',\\ &= 0, \quad \text{otherwise,}\end{aligned}$$

and $L(\mathbf{d}^{(1)},\dots,\mathbf{d}^{(\mu)})=C$, $\{\mu=(r-1)(s-1)\}$.

Then if $d(\tilde{p},\tilde{q})=\{\sum(\mathbf{d}^{(\nu)},\boldsymbol{\alpha})^2\}^{\frac{1}{2}}$, $d(\tilde{p},\tilde{q})$ is a metric on the equivalence classes of $\overset{CR}{\sim}$, and the random variables $(\mathbf{d}^{(\nu)},\Theta)$ are, asymptotically, independent normal, with unit variances, and means $(\mathbf{d}^{(\nu)},\mathbf{a})$, where $a_{ij}=\log\nu_{ij}$, $1\leqslant i\leqslant r$, $1\leqslant j\leqslant s$. Hence the *asymptotic posterior distribution of* $\{d(\Phi,\tilde{\imath})\}^2$ *is non-central* χ^2, *with degrees of freedom* $(r-1)(s-1)$ *and parameter of non-centrality* $\sum(\mathbf{d}^{(\nu)},\mathbf{a})^2$. If $\nu_{ij}=n_{ij}$, $1\leqslant i\leqslant r$, $1\leqslant j\leqslant s$, the quantity $\sum(\mathbf{d}^{(\nu)},\mathbf{a})^2$ is what Lindley calls $\mathbf{n}'\mathbf{A}^{-1}\mathbf{n}$, which he refers to $\chi^2_{(r-1)(s-1)}$ to test for association of rows and columns of the contingency table. In this case the metric $d(\tilde{p},\tilde{q})$ chosen to measure the association depends, through the use of $\langle\cdot,\cdot\rangle$, on the particular observations made. However, as Lindley remarks, the quantity $\mathbf{n}'\mathbf{A}^{-1}\mathbf{n}$ is independent of the original cross-ratios chosen; in this notation this means that $\sum(\mathbf{d}^{(\nu)},\mathbf{a})^2$ is the same for any basis $(\mathbf{d}^{(\nu)})$ of C that is orthonormal with respect to $\langle\cdot,\cdot\rangle$.

Acknowledgement

I should like to thank my research supervisor Mr A. D. McLaren for much helpful advice and encouragement, and the Medical Research Council for financial support. I am also grateful to the Cambridge University Mathematical Laboratory for computing time on TITAN, and to the referee for his comments.

References

ALI, S. M. and SILVEY, S. D. (1965). Association between random variables and the dispersion of a Radon–Nikodym derivative. *J. R. Statist. Soc.* B, **27**, 100–107.

BIRCH, M. W. (1963). Maximum likelihood in three-way contingency tables. *J. R. Statist. Soc.* B, **25**, 220–233.

BLOCH, D. A. and WATSON, G. S. (1967). A Bayesian study of the multinomial distribution. *Ann. Math. Statist.*, **38**, 1423–1435.

EDWARDS, A. W. F. (1963). The measure of association in a 2×2 table. *J. R. Statist. Soc.* A, **126**, 109–114.

GOOD, I. J. (1965). *The Estimation of Probabilities: An Essay on Modern Bayesian Methods.* Research Monographs No. 30. Cambridge, Mass.: M.I.T. Press.

GOODMAN, L. A. and KRUSKAL, W. H. (1954). Measures of association for cross-classifications. *J. Amer. Statist. Ass.*, **49**, 732–764.

—— (1959). Measures of association for cross-classification, II: Further discussion and references. *J. Amer. Statist. Ass.*, **54**, 132–163.

LEHMANN, E. L. (1966). Some concepts of dependence. *Ann. Math. Statist.*, **37**, 1137–1153.

LINDLEY, D. V. (1964). The Bayesian analysis of contingency tables. *Ann. Math. Statist.*, **35**, 1622–1643.

PLACKETT, R. L. (1962). A note on interactions in contingency tables. *J. R. Statist. Soc.* B, **24**, 162–166.

RÉNYI, A. (1959). On measures of dependence. *Acta Math. Acad. Sci., Hung.*, **10**, 441–451.

SINKHORN, R. (1967). Diagonal equivalence to matrices with prescribed row and column sums. *Amer. Math. Mon.*, **74**, 402–405.