A Distance Representation of the Quasi-Symmetry Model and Related Distance Models

Mark de Rooij[1] and Willem J. Heiser[2]

[1] Department of Psychology, Leiden University *rooijm@fsw.leidenuniv.nl*
[2] Department of Psychology, Leiden University *heiser@fsw.leidenuniv.nl*

Summary. We propose a complete distance representation for the quasi-symmetry model for the analysis of square contingency tables. Complete in the sense that both interaction and main effects will be represented in a single distance model. Distances represent a departure from the maximum frequency in the contingency table. The model is explained in some detail and applied to occupational mobility data. Finally, it is compared to existing multidimensional scaling models for asymmetric tables.

Key words. Asymmetry, Contingency tables, Multidimensional scaling.

1 Square contingency tables

The analysis of square contingency tables often asks for symmetric or close to symmetric models. Basic loglinear modeling provides two models: the symmetry model and the quasi-symmetry model (see Agresti, 1990, Chapter 10). The quasi-symmetry model can be written as

$$\log(F_{ij}) = g + r_i + c_j + s_{ij}, \text{ with } s_{ij} = s_{ji}, \tag{1}$$

where g is a constant, r_i a row-effect ($i = 1, \ldots, I$), c_j a column-effect ($j = 1, \ldots, I$), s_{ij} the interaction effect which is symmetric, i.e. $s_{ij} = s_{ji}$, and F_{ij} is the resulting expected frequency. When we constrain the row- and column-effects to be the same, i.e. $r_i = c_i$ we obtain the symmetry model.

In the present paper we will discuss a distance representation of the quasi-symmetry model. First, the interaction effects will be transformed to Euclidean distances; second, the main effects will be rescaled to unique dimensions. The resulting model has an extremely simple interpretation in terms of distances: the smaller the distance between two categories the larger the transition frequency; the larger the distance the smaller the transition frequency. So, distances give a departure from the maximum frequency in the data.

When we have discussed our model and an application, we will discuss relationships with other distance models for asymmetric tables. The models to be discussed are the distance density model by Krumhansl (1978), the extended Euclidean model by Winsberg and Carroll (1989), the slide-vector model by Zielman and Heiser (1993), the wind model by Gower (1977), and

Table 1. Occupational Mobility data (Goodman, 1991)

	1	2	3	4	5	6	7
1	50	19	26	8	18	6	2
2	16	40	34	18	31	8	3
3	12	35	65	66	123	23	21
4	11	20	58	110	223	64	32
5	14	36	114	185	715	258	189
6	0	6	19	40	179	143	71
7	0	3	14	32	141	91	106

models by Okada and Imaizumi (1987), Weeks and Bentler (1982), and Saito (1991).

Before we discuss modeling we will show a data set here, to be used later in the application. It is an occupational mobility data set obtained from Goodman (1991), reproduced in Table 1. Both ways have seven occupational categories, here simply denoted by the numbers 1 to 7. From these data we have removed the diagonal, which is often done in the analysis of square contingency tables. The symmetry model does not fit the data ($X^2 = 50.6$, $G^2 = 54.0$, $df = 21$), but the quasi-symmetry model provides an adequate fit ($X^2 = 13.1$ and $G^2 = 15.6$ with $df = 15$).

2 Distance representation

2.1 Distance representation of the interaction

Multidimensional scaling models have been strongly developed in psychology and were initially used for the analysis of (dis)similarity judgments. Nowadays, these models have a much larger field of application. Scaling models have the virtue of a simple interpretation, since distances are encountered in every day life. In the present paper we will deal with the frequencies as similarity measures and so the distances must be a monotone decreasing function of the frequencies. In the quasi-symmetry model we will transform the interaction parameters s_{ij} to distances using the Gaussian transform (Shepard, 1958, p. 249; Nosofsky, 1985, p. 422). The model we will work with is defined by

$$\log(F_{ij}) = g + r_i + c_j - d_{ij}^2(\mathbf{X}), \qquad (2)$$

where the squared distance is defined as

$$d_{ij}^2(\mathbf{X}) = \sum_{p=1}^{P}(x_{ip} - x_{jp})^2. \qquad (3)$$

The $I \times P$ matrix $\mathbf{X}$ contains coordinates, x_{ip}, of the I categories on P dimensions. If $P = I - 1$ the number of parameters is the same as in the quasi-symmetry model, but of course dimensionality restrictions can be imposed. De Rooij and Heiser (submitted paper) coin model (2) the *symmetric distance association model* and present an algorithm to approximate observed frequencies (f_{ij}) with expected frequencies (F_{ij}) using a maximum likelihood function and assuming a Poisson sampling distribution.

2.2 Distance representation of the row- and column-effects

Besides the interaction, the main effects for rows and columns can also be transformed such that they can be incorporated in the distance representation. Therefore, we first make them all negative, i.e. we subtract a value from the row/column-effects such that the largest equals zero, and add these values to the constant g, that is

$$\begin{aligned}\log(F_{ij}) = g &+ \max_i(r_i) + \max_j(c_j) \\ &+r_i - \max_i(r_i) \\ &+c_j - \max_j(c_j) \\ &-d^2_{ij}(\mathbf{X}).\end{aligned} \tag{4}$$

Take the square root of the absolute values of the row parameters and denote them by u_i, i.e. $u_i = \sqrt{\mid (r_i - \max_i(r_i) \mid}$, and analogously for the column parameters to obtain v_j. We can add K unique dimensions for the rows and L unique dimensions for the columns to our graphical representation by defining $u_{ik} = u_i$ if $i = k$ otherwise $u_{ik} = 0$, and $v_{jl} = v_j$ if $j = l$ otherwise $v_{jl} = 0$. Our complete distance model can be written as

$$\begin{aligned}\log(F_{ij}) = g^* & \\ &- \sum_{k=1}^{K}(u_{ik} - u_{jk})^2 \\ &- \sum_{l=1}^{L}(v_{il} - v_{jl})^2 \\ &- \sum_{p=1}^{P}(x_{ip} - x_{jp})^2 \\ &= g^* - d^2_{ij}(\mathbf{X};\mathbf{U};\mathbf{V}),\end{aligned} \tag{5}$$

where $g^* = g + \max_i(r_i) + \max_j(c_j)$ and denotes the maximum frequency from which distances are subtracted. Notice that we still have the same expected frequencies as in model (2), we just changed the identification constraints.

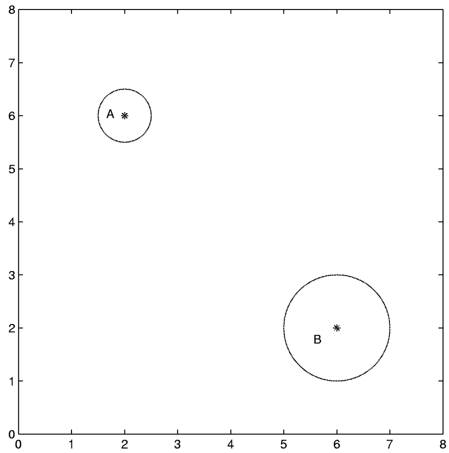

Fig. 1. Configuration with unique dimensions as circles

2.3 Graphical representation in low dimensionality

A disadvantage of the formulation as it is up to this point, is the huge number of dimensions. Suppose we have only two dimensions for the interaction, then our graphical representation has $2 + 2(I - 1)$ dimensions, and basically our graphical representation is lost. We can, however, draw the unique contributions in our two-dimensional representation and maintain the distance interpretation. If we draw circles around each point with radius u_{ik} for the point being a row category, and a circle with radius v_{jl} for the point being a column category, we obtain a novel interpretation of this distance model (see Figure 1, where only the row circle for category A is shown and the column circle for category B).

Since the model is defined in squared distances, by repeated use of Pythagoras theorem we can obtain the complete distance. This is shown in Figure 2. In the first step, we draw a radius (Bb) orthogonal to line AB (Left figure) and by Pythagoras we have that the squared length of the line Ab is equal to $AB^2 + Bb^2$. Then we repeat this (right figure) and draw the radius Aa orthogonal on Ab, the square of the length of ab is the complete distance, i.e. the deviance from the maximum frequency.

2.4 Asymmetry

In our model, each category is represented by one point and two circles, a circle for the row-effect and a circle for the column effect. The distance between two points is symmetric, i.e. $d_{ij}(\mathbf{X}) = d_{ji}(\mathbf{X})$. However, the complete distance between the two categories is not symmetric, i.e. $d_{ij}(\mathbf{X};\mathbf{U};\mathbf{V}) \neq d_{ji}(\mathbf{X};\mathbf{U};\mathbf{V})$. The asymmetry in the data is represented by the radii of the

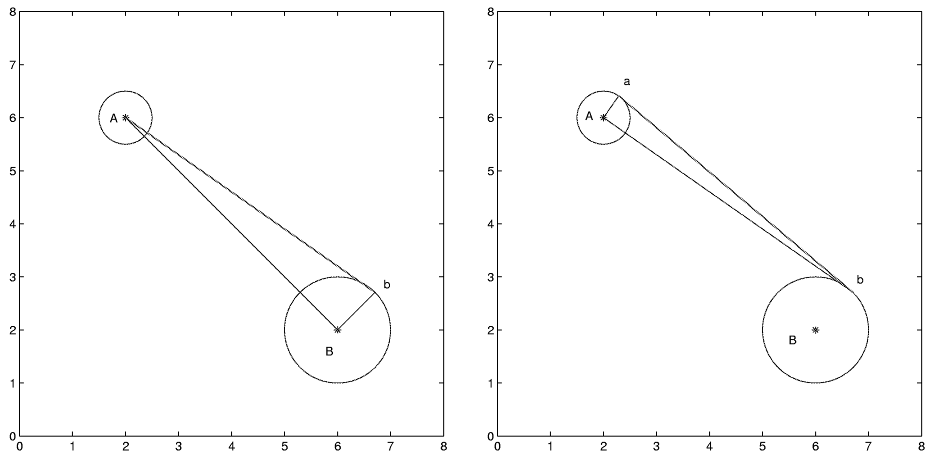

Fig. 2. Obtaining the complete distance

circles. See Figure 3 where point A has a dotted circle representing the row-effect and a solid circle representing the column-effect, for point B the row and column effect are equal, i.e. the same circle. We obtain complete distances as discussed above and we see that the complete distance from A to B (from dotted circle around A to circle around B) is smaller than the complete distance from B to A (from circle around B to solid circle around A). Since the expected frequencies are equal to the maximum frequency minus the complete distance, we have that F_{AB} is larger than F_{BA}.

2.5 Mass

The radius of the circle is inversely related to the mass of the corresponding category. So, in Figure 3, the mass of A being a row point (dotted circle) is larger than for A being a column point (solid circle). The advantage of the inverse relationship is outlined above, i.e. the advantage is having a distance representation. The disadvantage is clear and a warning is on its place. A natural interpretation of the circles in terms of mass is that the larger the circle the larger the mass. This is not true in the representation given above.

3 Application

Applying our model to the occupational mobility data we find that the two-dimensional model fits well, $X^2 = 16.0$ and $G^2 = 18.5$ with $df = 18$. The fit hardly decreases but the number of degrees of freedom increases from 15 to 18 compared to the quasi-symmetry model, so the model is more parsimonious. The graphical representation is given in Figure 4.

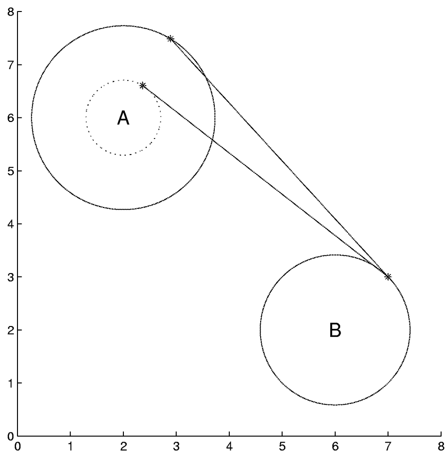

Fig. 3. Asymmetry in the graphical representation

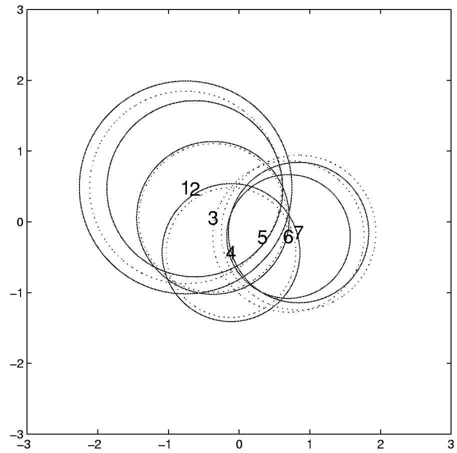

Fig. 4. Solution of occupational Mobility data

First, the points lie neatly ordered from 1 till 7. Since this was also found by Goodman (1991) using the $RC(M)$-association model, it appears to be some status ordering, where transitions occur more often between two adjacent categories than between categories far apart. The conclusion might be that going up or down the social status order goes, in general, smoothly step by step.

The unique contributions are drawn for the rows by dotted circles and for the columns by solid circles. Category 5 has for both the rows and the columns no circle, i.e. category five has largest mass by both sons and fathers. We see that the dotted and solid circles are basically the same, there is few asymmetry in the data or the masses for fathers and sons are about equal. For categories 1, 3, and 4 the solid circles are larger than the dotted circles; For categories 6 and 7 the dotted circles are larger than the solid circles. Since transitions always go from row (dotted) to column (solid) we have in general that a transition up the social status ladder occurs more frequently than a transition down the social ladder. For example, the complete distance from category 7 to 1 is larger than the complete distance from 1 to 7. So, the expected transition frequency from 1 to 7 is larger than that from 7 to 1. This is also true in the observed data.

4 Related distance models

Before comparing our model to related distance models we will first write our model in matrix terminology. Model 2 can be written as

$$\log(\mathbf{F}) = g\mathbf{1}\mathbf{1}' + \mathbf{r}\mathbf{1}' + \mathbf{1}\mathbf{c}' - \mathbf{D}^2(\mathbf{X}). \tag{6}$$

Table 2. Related distance models and their matrix expression

Model	Matrix expression
Distance Density Model	$\phi(\mathbf{\Delta}) = \mathbf{D} + w_1\mathbf{m}\mathbf{1}' + w_2\mathbf{1}\mathbf{m}'$
Extended Euclidean Model	$\phi(\mathbf{\Delta}) = \mathbf{D}^2 + \mathbf{m}\mathbf{1}' + \mathbf{1}\mathbf{m}'$
Slide-Vector Model	$\phi(\mathbf{\Delta}) = \mathbf{D}_z^2 + 2(\mathbf{X}\mathbf{z}\mathbf{1}' - \mathbf{1}\mathbf{z}'\mathbf{X}')$ $\phi(\mathbf{\Delta}) = \mathbf{D}_z^2 + \mathbf{n}\mathbf{1}' - \mathbf{1}\mathbf{n}'$
Wind Model	$\phi(\mathbf{\Delta}) = \mathbf{D} + \mathbf{n}\mathbf{1}' - \mathbf{1}\mathbf{n}'$
Okada & Imaizumi Model	$\phi(\mathbf{\Delta}) = \mathbf{D} - \mathbf{n}\mathbf{1}' + \mathbf{1}\mathbf{n}'$
Weeks & Bentler Model	$\phi(\mathbf{\Delta}) = b\mathbf{D} + k(\mathbf{1}\mathbf{1}' - \mathbf{I}) + \mathbf{n}\mathbf{1}' - \mathbf{1}\mathbf{n}'$
Saito Model	$\phi(\mathbf{\Delta}) = \mathbf{D} + \mathbf{M} + \mathbf{n}\mathbf{1}' - \mathbf{1}\mathbf{n}'$

If we define

$$\begin{aligned}\mathbf{m} &= (\mathbf{r} + \mathbf{c})/2 \\ \mathbf{n} &= (\mathbf{r} - \mathbf{c})/2,\end{aligned} \tag{7}$$

and

$$\begin{aligned}\mathbf{M} &= \mathbf{m}\mathbf{1}' + \mathbf{1}\mathbf{m}' + g\mathbf{1}\mathbf{1}' \\ \mathbf{N} &= \mathbf{n}\mathbf{1}' - \mathbf{1}\mathbf{n}',\end{aligned} \tag{8}$$

then $\mathbf{M}$ is symmetric and $\mathbf{N}$ is skew-symmetric. The idea to decompose the parameters of an asymmetric model into a symmetric and a skew-symmetric part was first fully exploited by Zielman and Heiser (1996).

Model 2 is then:

$$\log(\mathbf{F}) = \mathbf{M} + \mathbf{N} - \mathbf{D}^2(\mathbf{X}). \tag{9}$$

In multidimensional scaling we often work with dissimilarities to which we fit distances. The frequencies are similarity measures, and so the negation of frequencies are dissimilarities up to a constant. We can then write

$$-\log(\mathbf{F}) = \phi(\mathbf{\Delta}) = \mathbf{D}^2(\mathbf{X}) - \mathbf{M} - \mathbf{N}, \tag{10}$$

where of course the sign on $\mathbf{M}$ and $\mathbf{N}$ is arbitrary.

In Table 2 we give different distance models and their matrix expressions. The models discussed are Krumhansl's (1978) distance density model; the extended Euclidean model by Winsberg and Carroll (1989); the slide-vector model by Zielman and Heiser (1993); the wind model by Gower (1977); and models by Okada and Imaizumi (1987), Weeks and Bentler (1982, and Saito (1991). An important difference is that our model is estimated by maximizing a likelihood function where all these models are estimated by minimizing a least squares function. Comparing the models, we see that all models use either the vector $\mathbf{m}$ as defined above or the vector $\mathbf{n}$ to model skew-symmetry. Only Saito's model uses both terms. In the distance density model the asymmetry is defined by two weights, one for the rows and one for the columns.

The extended Euclidean model is a symmetric model extending the standard Euclidean part with unique dimensions. The last three models are distance models for asymmetric tables, where we have a distance part and some departure from the symmetric distance by a skew-symmetric term. Note that in the slide vector model we have $\mathbf{D}_z^2$ denoting that it is a distance matrix with a constant added to all cells. For a further discussion of these relationships see Gower (2000).

References

Agresti A (1990) Categorical data analysis. John Wiley and sons, New York

De Rooij M, Heiser WJ (2002, submitted) Graphical representations and odds ratios in a distance-association model for the analysis of cross-classified data

Goodman LA (1991) Measures, models, and graphical displays in the analysis of cross-classified data. Journal of the American Statistical Association, 86: 1085-1111

Gower JC (1977) The analysis of asymmetry and orthogonality. In: Barra JR, Brodeau F, Romer G, Van Cutsem B (eds) Recent developments in statistics. North-Holland, Amsterdam, pp 109-123

Gower JC (2000) Rank-one and rank-two departures from symmetry. Computational statistics and data analysis 33: 177-188

Krumhansl C (1978) Concerning the applicability of geometric models to similarity data: the interrelationship between similarity and spatial density. Psychological Review 85: 445-463

Nosofsky RM (1985) Overall similarity and the identification of seperable dimension stimuli: a choice model analysis. Perception and Psychophysics 38: 415-432

Okada A, Imaizumi T (1987) Nonmetric multidimensional scaling of asymmetric proximities. Behaviormetrika 21: 81-96

Saito T (1991) Analysis of asymmetric proximity matrix by a model of distance and additive terms. Behaviormetrika 29: 45-60

Shepard RN (1958) Stimulus and response generalization: deduction of the generalization gradient from a trace model. Psychological Review 65: 242-256

Weeks DG, Bentler PM (1982) Restricted multidimensional scaling models for asymmetic proximities. Psychometrika 47: 201-208

Winsberg S, Carroll JD (1989) A quasi-nonmetric method for multidimensional scaling via an extended Euclidean model. Psychometrika 54: 217-229

Zielman B, Heiser WJ (1993) The analysis of asymmetry by a slide-vector. Psychometrika 58: 101-114

Zielman B, Heiser WJ (1996) Models for asymmetric proximities. British Journal of Mathematical and Statistical Psychology 49: 127-146.