Statistica Neerlandica (2001) Vol. 55, nr. 2, pp. 157–181

Distance association models for the analysis of repeated transition frequency tables

Mark de Rooij*

Department of Psychology, P.O. Box 9555, 2300 RB Leiden, The Netherlands

The present paper is concerned with the analysis of repeated transition frequency tables, for example, occupational mobility data measured in different cohorts. The association present in such a table will be modeled by a distance in Euclidean space. A large distance corresponds to a small association; a small distance corresponds to a large association. A more direct interpretation is that more transitions occur between categories that are close together in a social space. It is assumed that the same social structure (space) exists for the different slices (cohorts/time points) of a table, but that the dimensions of this space are weighted for the different slices, i.e., for each slice the dimensions are stretched or squeezed. We will propose a model, discuss an algorithm to obtain maximum likelihood estimates and apply the model to an empirical data set.

Key Words and Phrases: three-way table; weighted Euclidean distance; Gaussian transformation; log–linear model; maximum likelihood.

1 Introduction

The present paper considers the analysis of multiple transition frequency tables, that is transition frequency tables measured at different time-points, in different countries, or in different cohorts. For the analysis of a single transition frequency matrix DE ROOIJ and HEISER (2000) presented distance association models. In these models the association parameters of a log–linear model are transformed to distances in Multidimensional Euclidean space. The idea of using distances for square contingency tables is quite old: SHEPARD (1957) transformed association parameters of a multiplicative model to distances in the context of stimulus recognition data and stimulus generalization data; GOODMAN (1972) and HABERMAN (1974) describe models for square contingency tables, where they use one-dimensional distances, and applied these to occupational mobility data. In the field of stimulus recognition data the early work of Shepard is further developed (see for example a book edited by ASHBY, 1992). DE ROOIJ and HEISER (2000) generalized the work of GOODMAN and HABER-

* rooijm@fsw.leidenuniv.nl

MAN to the multidimensional case. Here we will generalize the symmetric distance association model proposed by DE ROOIJ and HEISER (2000) to the case of three-way tables. We assume one common association space, where a small distance corresponds to a large association, and a large distance to a small association. This common space is weighted for the different time-points/countries/cohorts dimension wise. In other words, every time-point/country/cohort obtains a weight for each dimension that shrinks (weight < 1) or stretches (weight > 1) the corresponding dimension. The distance part of the model to be proposed, is equal to the weighted Euclidean distance, maybe better known as the INDSCAL-model (CARROLL and CHANG, 1970). As in the distance association models proposed by DE ROOIJ and HEISER (2000), association parameters of a log–linear model are transformed to Euclidean distances. One major advantage of this reparametrization is the fast and simple interpretation of the final result. A second advantage is a significant reduction in the number of parameters, especially for large tables.

Earlier papers that describe the weighted Euclidean model in a longitudinal context are, among others, MCCAIN (1998), WHITE and MCCAIN (1998), RIKKEN, KIERS and VOS (1995), and ROTH *et al*. (1992). In none of these papers, the parameters of the formulated model are estimated using maximum likelihood theory, but all use the least squares methodology as in the original formulation by CARROLL and CHANG (1970). CARROLL and GREEN (1997) discuss the use of scaling methods in marketing research. In order to increase the practical utility of scaling methods in marketing research, they suggest developing scaling methods under maximum likelihood using appropriate distributional assumptions. Since for repeated contingency tables the distributional properties are well known, we will use these properties and estimate the model parameters using maximum likelihood. RAMSAY (1977, 1982) also proposed estimating the model parameters of the weighted Euclidean model by maximum likelihood methods. He, however, assumed a normal or log–normal distribution for his dissimilarity data. For contingency tables the usual distributional assumptions are either a Poisson or (product) multinomial distribution. We will utilize these latter distributional assumptions in this paper.

Before we go into detail about the model, let us have a look at a data set that will be used to illustrate the proposed methodology. Table 1 is obtained from GANZEBOOM and LUIJKX (1995) and considers occupational mobility in The Netherlands in the period 1970–1993. Five sub-periods are considered, 1970–1974, 1975–1979, 1980–1984, 1985–1989, and 1990–1993. Ten occupational categories were formed using the EGP-scheme (ERIKSON, GOLDTHORPE and PORTOCARERO, 1979). The following categories are distinguished: (1) Large proprietors, higher professionals and managers; (2) Lower professionals and managers; (3) Routine non manual workers; (4a) Small proprietors with employees; (4b) Small proprietors without employees; (5) Lower grade technicians and manual supervisors; (6) Skilled manual workers; (7a) Unskilled and semi-skilled manual workers; (9c) Self employed farmers; and (7b) (Unskilled) agricultural workers. In these kind of data an important question is: ‘What do the patterns of intergenerational mobility look like, i.e., what is

Table 1. Intergenerational Occupational Mobility Table of men in The Netherlands, 1970–1993. Obtained from GANZEBOOM and LUIJKX (1995). For a description of the categories see text.

father	*son* (1)	(2)	(3)	(4a)	(4b)	(5)	(6)	(7a)	(9c)	(7b)
					1970–1974					
(1)	40	30	24	0	5	0	14	7	1	1
(2)	60	91	40	3	4	4	34	9	3	1
(3)	27	43	41	2	2	1	19	10	0	0
(4a)	3	18	8	18	14	0	12	13	0	0
(4b)	16	51	45	18	36	4	36	27	4	3
(5)	2	5	1	0	0	0	4	2	0	0
(6)	25	41	49	5	6	6	113	49	1	2
(7a)	21	32	31	5	9	6	98	90	4	11
(9c)	15	37	26	6	15	2	45	38	116	21
(7b)	3	7	15	1	6	2	29	46	8	23
					1975–1979					
(1)	97	63	20	7	4	2	22	14	1	1
(2)	91	139	54	4	21	11	32	20	2	2
(3)	47	78	72	4	7	9	45	18	0	2
(4a)	35	34	31	23	17	9	24	15	2	2
(4b)	53	59	57	28	43	11	55	40	7	5
(5)	16	26	24	3	3	12	24	19	0	0
(6)	36	95	83	6	14	40	211	74	3	7
(7a)	30	80	86	11	16	38	163	139	5	13
(9c)	31	69	32	17	17	20	65	60	166	33
(7b)	15	17	15	2	8	9	37	43	11	22
					1980–1984					
(1)	66	78	48	2	2	7	21	19	0	3
(2)	67	109	74	4	9	7	20	23	0	2
(3)	38	84	69	2	4	8	29	18	2	1
(4a)	17	23	23	9	9	5	23	11	1	1
(4b)	35	47	49	22	25	3	50	29	7	1
(5)	16	26	22	1	2	12	26	11	0	1
(6)	42	105	105	6	12	33	158	101	1	5
(7a)	24	65	80	5	16	23	130	114	1	9
(9c)	30	43	42	9	8	9	53	63	92	16
(7b)	9	11	14	0	2	7	31	27	4	16
					1985–1989					
(1)	84	89	50	5	3	6	20	24	2	1
(2)	67	179	83	7	10	17	48	44	5	2
(3)	65	128	115	8	7	14	69	62	2	4
(4a)	30	39	38	26	6	2	29	30	3	2
(4b)	47	53	48	19	24	7	46	52	2	1
(5)	18	35	22	0	1	11	24	29	0	2
(6)	76	132	129	19	12	40	234	170	6	5
(7a)	71	117	110	9	13	28	193	220	3	21
(9c)	43	78	53	9	7	9	74	69	119	21
(7b)	9	16	29	1	6	2	28	47	5	12

continued overleaf

Table 1. (*continued*)

father	*son* (1)	(2)	(3)	(4a)	(4b)	(5)	(6)	(7a)	(9c)	(7b)
					1990–1993					
(1)	52	94	43	4	7	5	11	17	2	1
(2)	50	121	42	3	6	9	37	26	2	5
(3)	36	65	40	7	4	5	23	20	3	1
(4a)	21	31	17	18	2	1	9	12	0	0
(4b)	8	23	11	6	1	2	14	12	0	1
(5)	14	19	11	2	2	6	8	10	0	1
(6)	32	76	48	5	1	12	80	65	2	1
(7a)	21	45	39	3	7	11	63	59	5	3
(9c)	23	52	14	3	3	7	33	32	61	10
(7b)	5	10	10	1	0	3	19	13	2	2

the social distance between occupational categories?' (GANZEBOOM and LUIJKX, 1995, p. 14). A second important question is with regard to the historical changes in intergenerational mobility: 'Are occupational categories getting closer or further away from each other?' (GANZEBOOM and LUIJKX, 1995, p. 15). We will show that our distance models developed in this paper are highly suitable for answering these questions.

The remainder of this paper is organized as follows. The next section describes the model and some properties of the model. The third section discusses an algorithm to obtain solutions to the model. In Section 4 we apply the models to the data described above, and provide a comparison with existing models. We conclude this paper with a discussion of the obtained results.

2 The Model

As stated in the introduction, we will transform association parameters of a log–linear model to distances in Euclidean space. First, consider the full log–linear model for a three-way contingency table

$$\log(\pi_{ijk}) = \lambda + \lambda_i^R + \lambda_j^C + \lambda_k^P + \lambda_{ij}^{RC} + \lambda_{ik}^{RP} + \lambda_{jk}^{CP} + \lambda_{ijk}^{RCP},$$

where π_{ijk} is the expected frequency for cell ijk, $i = 1, \ldots, I$, $j = 1, \ldots, J$, and $k = 1, \ldots, K$. Furthermore, λ denotes the general mean, λ_i^R, λ_j^C, and λ_k^P, denote the main effects of the three variables, and λ_{ij}^{RC}, λ_{ik}^{RP}, λ_{jk}^{CP}, and λ_{ijk}^{RCP}, denote respectively the two-way and three-way association terms. Many authors have studied these models, we refer to BISHOP, FIENBERG and HOLLAND (1975), FIENBERG (1980), and AGRESTI (1990).

In the present paper we consider a special case, that is the case where $I = J$, each slice is square. We will transform all association parameters (i.e., λ_{ij}^{RC}, λ_{ik}^{RP}, λ_{jk}^{CP}, and λ_{ijk}^{RCP}) to a distance in Euclidean space d_{ijk} by the Gaussian transformation (see

SHEPARD, 1957, 1987, NOSOFSKY, 1985; DE ROOIJ and HEISER (2000)). The Gaussian transformation is a monotonically decreasing function, such that a large distance corresponds to a small value for the association, and a small distance corresponds to a large value for the association. We do not distinguish between first-order association and second-order association, but all association present in the data is modeled by distances in Euclidean space. In other words, the distances represent the departure from independence.

We will use the weighted Euclidean distance (CARROLL and CHANG, 1970) for modeling the association. The weighted Euclidean distance in squared form is given by

$$d_{ijk}^2(\mathbf{X}; \mathbf{W}_k) = \sum_m w_{km}^2 (x_{im} - x_{jm})^2$$

$$= \sum_m (w_{km} x_{im} - w_{km} x_{jm})^2,$$

where x_{im} is the coordinate for category i on the m-th dimension ($m = 1, \ldots, M$), which are captured in a $I \times M$-matrix $\mathbf{X}$, and w_{km} is the weight for slice k for dimension m, captured in a diagonal matrix $\mathbf{W}_k$. If the weights w_{km} are equal to one, the weighted Euclidean distance simplifies to the Euclidean distance between points i and j. Transforming all association parameters to the weighted Euclidean distance using the Gaussian transformation we obtain the following model:

$$\log(\pi_{ijk}) = \lambda + \lambda_i^R + \lambda_j^C + \lambda_k^P - d_{ijk}^2(\mathbf{X}; \mathbf{W}_k). \quad (1)$$

In square contingency tables the frequencies on the diagonal are often relatively large compared with the off-diagonal frequencies. In longitudinal research, these large frequencies come about since people in general do not change. If we try to fit a model to a square contingency table, the model often does not fit because of the large discrepancies of the diagonal cells. Additional parameters can be incorporated into the model to adjust for these diagonal cells. A first way to adjust the model is to fit parameters to the diagonal in every slice k. The model then becomes

$$\log(\pi_{ijk}) = \lambda + \lambda_i^R + \lambda_j^C + \lambda_k^P - d_{ijk}^2(\mathbf{X}; \mathbf{W}_k) + \delta_{ij}\epsilon_i^k, \quad (2)$$

where δ_{ij} is the Kronecker delta, and ϵ_i^k are the parameters that fit the diagonal expected frequencies to the corresponding observed frequencies in slice k. In this case $I \times K$ additional parameters are fitted, and the expected frequencies on the diagonal of each slice k are set equal to the observed frequencies of that slice. Another option is to fit only one set of parameters to the diagonal, i.e., restrict all ϵ_i^k to be equal for different k. The model then can be written as

$$\log(\pi_{ijk}) = \lambda + \lambda_i^R + \lambda_j^C + \lambda_k^P - d_{ijk}^2(\mathbf{X}; \mathbf{W}_k) + \delta_{ij}\epsilon_i, \quad (3)$$

and only I additional parameters are needed. Since the distance from a point towards itself is zero, the distances and the parameters for the diagonal are not confounded.

Returning to model (1) for the moment, and using the definition of the weighted Euclidean distance, we obtain the following:

$$\begin{aligned}\log(\pi_{ijk}) &= \lambda + \lambda_i^R + \lambda_j^C + \lambda_k^P - \sum_m w_{km}^2 (x_{im} - x_{jm})^2 \\ &= \lambda + \lambda_i^R + \lambda_j^C + \lambda_k^P - \sum_m w_{km}^2 x_{im}^2 \\ &\quad - \sum_m w_{km}^2 x_{jm}^2 + 2 \sum_m w_{km}^2 x_{im} x_{jm} \\ &= \lambda + \lambda_i^R + \lambda_j^C + \lambda_k^P - u_{ik} - u_{jk} + 2 \sum_m w_{km}^2 x_{im} x_{jm}. \end{aligned} \tag{4}$$

From this reformulation we see that it would not make much sense to include first-order interaction terms into model (1), since then these terms would be confounded with the u-terms in the formulation above, and the interpretation gets troublesome.

To enhance interpretation of our distance model, consider the following: if the dimensionality of our model is zero, the model becomes the model of *mutual independence* (AGRESTI, 1990, pp. 138–139). If all coordinates for every dimension are equal, the model also reduces to the model of mutual independence. If the weights are equal to each other, i.e., $w_{km} = 1$ for all k and m, the rows and the columns are *jointly independent* of the planes (slices). The two-way association effects (the u-terms in (4)) are equal for the association between the rows and the planes and the columns and the planes, because of the equality of row and column coordinates.

DE ROOIJ and HEISER (2000) show their distance association models are reparametrization of the $RC(M)$-association model proposed by GOODMAN (1979, 1985). A generalization of the $RC(M)$-association model for multiple sets is given by BECKER and CLOGG (1989). They proposed multiple group association models, for the analysis of a set of cross-classifications. The model of Becker and Clogg is not restricted to repeated square contingency tables. The model they propose has the following form

$$\pi_{ijk} = \alpha_{i(k)} \beta_{j(k)} \exp\left\{ \sum_{m=1}^{M_k} \phi_{m(k)} \mu_{im(k)} \nu_{jm(k)} \right\}. \tag{5}$$

The α and β terms are the usual first-order association terms (the main effects are implied), the ϕ parameters are called intrinsic association parameters and the μ and ν-parameters are row and column scores, respectively. The row and column scores are restricted to have zero mean and unit variance. Becker and Clogg propose restricting parameters to be equal for different slices k, to obtain homogeneous models. Each of the terms in the exponent part of the model can be restricted. If all terms are independent of k, the model is equal to the model of no three-way association. If $M_k = 0$ for all k the model is equal to the model of conditional independence of the row and column variable. If $M_k = M^* = \min(I-1, J-1)$ for all k the model is equal to the saturated model.

Comparing this model with our distance association model we find that the multiple group association model has first-order interaction parameters for the association between the row variable and the plane variable, and the column variable and the plane variable. In our model these are confounded in the distance term, i.e., the u-terms in (4). The restrictions placed on the solution are not the same in our model and as in the multiple group association model: we restrict the weights to have a sum equal to one in each dimension, where the row and column scores are restricted in the multiple group association model; in our distance association model the coordinates for the rows and columns are restricted to be equal, whereas in the multiple group association model the scores for the rows and columns are not related.

2.1 *Degrees of freedom*

It is well known that the maximum rank of a three-way table or a higher-way table is unknown (KRUSKAL, 1977, 1989; TEN BERGE *et al*., 1988; TEN BERGE, 1991; SICILIANO and MOOIJAART, 1997). Therefore it is difficult to assess the degrees of freedom for a particular model. The number of degrees of freedom for model (1) in this paper is obtained as follows: for the λ-parameters we subtract the usual number, i.e., the number of categories minus one. It is well known that the weighted Euclidean model has no rotational freedom, so the number of parameters for the distances are $(I-1) \times M$, since the solution can be centered without loss of generality. For the weights we have to estimate $K \times M$ parameters, which are constrained to have mean equal to one in every dimension. The number of independent parameters for model (1) is equal to $2(I-1) + (K-1) + (I-1) \times M + (K-1) \times M$. For models with diagonal terms, I or $K \times I$ have to be added to this number of independent parameters. We will subtract this number of independent parameters from the number of cells minus one, as is most often done in the analysis of three-way contingency tables. The total number of degrees of freedom the model equation (1) then becomes $IJK - I - J - K + 2 - (I + K - 2) \times M$.

2.2 *Odds ratios*

As is shown by DE ROOIJ (2000) the proposed model (1) does represent three-way association as measured by the second-order odds-ratio. We will briefly discuss the findings here. The second-order odds-ratio ($\theta_{ii'jj'kk'}$), also called the ratio of odds ratios for a $2 \times 2 \times 2$-subtable, is defined by

$$\theta_{ii'jj'kk'} = \frac{\pi_{ijk}\pi_{i'j'k}\pi_{i'jk'}\pi_{ij'k'}}{\pi_{i'j'k'}\pi_{i'jk}\pi_{ij'k}\pi_{ijk'}}.$$

This second-order odds ratio is a measure of three-way association, not dependent on the marginal frequencies of a table, which is an important property of a measure of association. Developing the logarithm of the second-order odds ratio under our model we obtain

$$\log(\theta_{ii'jj'kk'}) = 2\sum_m (w_{km}^2 - w_{k'm}^2) \times (x_{im}x_{jm} + x_{i'm}x_{j'm} - x_{i'm}x_{jm} - x_{im}x_{j'm})$$

$$= 2\sum_m (w_{km}^2 - w_{k'm}^2) \times (x_{im} - x_{i'm})(x_{jm} - x_{j'm}).$$

The logarithm of the second-order odds ratio equals zero if and only if either $w_{km} = w_{k'm}$, for all m, or $x_{im} = x_{i'm}$ for all m, or $x_{jm} = x_{j'm}$ for all m.

The conditional odds ratio is a measure of two-way association in one slice (k) of our three-way table, and is defined by

$$\theta_{ii'jj'|k} = \frac{\pi_{ijk}\pi_{i'j'k}}{\pi_{i'jk}\pi_{ij'k}}.$$

Under our model the conditional odds ratio is equal to

$$\log(\theta_{ii'jj'|k}) = \sum_m \big[w_{km}^2(x_{i'm} - x_{jm}) + w_{km}^2(x_{im} - x_{j'm})$$

$$-w_{km}^2(x_{im} - x_{jm}) - w_{km}^2(x_{i'm} - x_{j'm})\big]$$

$$= 2\sum_m w_{km}^2(x_{im}x_{jm} + x_{i'm}x_{j'm} - x_{i'm}x_{jm} - x_{im}x_{j'm})$$

$$= 2\sum_m w_{km}^2(x_{im} - x_{i'm})(x_{jm} - x_{j'm}).$$

This is the same form as derived by DE ROOIJ and HEISER (2000) for a two-way table (i.e., only one source) when we set the weight for each source k on each dimension equal to one (i.e., equal dimension weights). If we define y_{im}^k by $y_{im}^k = w_{km}x_{im}$ and collect these in a matrix $\mathbf{Y}^k$, we can rewrite the conditional odds ratio for slice k as

$$\log(\theta_{ii'jj'|k}) = d_{i'j}^2(\mathbf{Y}^k) + d_{ij'}^2(\mathbf{Y}^k) - d_{ij}^2(\mathbf{Y}^k) - d_{i'j'}^2(\mathbf{Y}^k),$$

so the conditional odds ratio can be fully expressed in terms of distances.

For the models including terms for the diagonal cells (models (3) and (2)) the expressions for the second-order odds ratio and the conditional odds ratio should be adapted to include these parameters.

3 ML-estimation and fit

3.1 *An elementary Newton algorithm*

To find a solution to the model we make use of an elementary Newton algorithm. In general an update for a parameter θ_i, is given by

$$\theta_i^{t+1} = \theta_i^t - h(\theta_i)^{-1}q(\theta_i),$$

where $h(\theta_i)$ and $q(\theta_i)$ are the second and first derivative of the likelihood function with respect to parameter θ_i, respectively.

The log-likelihood function for the most general model (2) can be written and developed as

$$\begin{aligned}
L = {} & f_{+++}\lambda + \sum_i f_{i++}\lambda_i^R + \sum_j f_{+j+}\lambda_j^C + \sum_k f_{++k}\lambda_k^P \\
& - \sum_{ijk} f_{ijk} d_{ijk}^2(\mathbf{X};\, \mathbf{W}_k) + \sum_{ijk} f_{ijk}\delta_{ij}\epsilon_i^k - \sum_{ijk} \\
& \times \exp\left\{\lambda + \lambda_i^R + \lambda_j^C + \lambda_k^P - d_{ijk}^2(\mathbf{X};\, \mathbf{W}_k) + \delta_{ij}\epsilon_i^k\right\} \\
= {} & f_{+++}\lambda + \sum_i f_{i++}\lambda_i^R + \sum_j f_{+j+}\lambda_j^C + \sum_k f_{++k}\lambda_k^P \\
& - \sum_{ijk} f_{ijk} \sum_m w_{km}^2(x_{im} - x_{jm})^2 + \sum_{ijk} f_{ijk}\delta_{ij}\epsilon_i^k - \sum_{ijk} \\
& \times \exp\left\{\lambda + \lambda_i^R + \lambda_j^C + \lambda_k^P - \sum_m w_{km}^2(x_{im} - x_{jm})^2 + \delta_{ij}\epsilon_i^k\right\}.
\end{aligned}$$

The first and second derivatives with respect to the marginal parameters and diagonal parameters are standard, see for example AGRESTI (1990, chapter 6). We will derive here the derivatives with respect to the coordinates and the weights. In the appendix we will provide an algorithm scheme.

First, for the coordinates x_{im}, the first derivative of the log-likelihood function is given by

$$q(x_{im}) = 2\sum_{jk}(\hat{\gamma}_{ijk} - g_{ijk}) \times w_{km}^2(x_{im} - x_{jm}), \tag{6}$$

where $g_{ijk} = f_{ijk} + f_{ijk}$ if $i \neq j$, else $g_{iik} = \sum_l g_{ilk}$, and $\hat{\gamma}_{ijk}$ is defined correspondingly on the expected frequencies. The second derivative is given by

$$h(x_{im}) = 2\sum_{jk}(\hat{\gamma}_{ijk} - g_{ijk})w_{km}^2 - 4\sum_{jk}(w_{km}^2 x_{im} - w_{km}^2 x_{jm})^2. \tag{7}$$

For the weights w_{km} the first derivative is given by

$$q(w_{km}) = 2\sum_{ij}(\hat{\pi}_{ijk} - f_{ijk}) \times w_{km}(x_{im} - x_{jm})^2, \tag{8}$$

where $\hat{\pi}_{ijk}$ is the expected frequency for cell ijk. The second derivative is given by

$$h(w_{km}) = 2\sum_{ij}(\hat{\pi}_{ijk} - f_{ijk}) \times (x_{im} - x_{jm})^2 - 4\sum_{ij}\hat{\pi}_{ijk}(w_{km}(x_{im} - x_{jm})^2). \tag{9}$$

Starting values

Here we will provide starting values for the algorithm; be aware that these starting values do not ensure that we will find a global optimum. Experience shows that the likelihood function has many local maxima. To get an idea whether we found a global maximum, multiple starts using different values should always be performed.

The λ-parameters: For these parameters we follow BECKER (1990). The starting value for the parameter corresponding to the general mean is $\lambda^0 = \sum_{ijk}\log f_{ijk}/I \times J \times K$. For the marginal parameters we set $\lambda_i^{R^0} = (\log f_{i++}/I) - \lambda^0$, $\lambda_j^{C^0} = (\log f_{+j+}/J) - \lambda^0$, and $\lambda_k^{P^0} = (\log f_{++k}/K) - \lambda^0$.

The coordinate-parameters: We can rewrite the likelihood function in matrix terms, we then obtain the following

$$\begin{aligned} L = {} & f_{+++}\lambda + \sum_i f_{i++}\lambda_i^R + \sum_j f_{+j+}\lambda_j^C + \sum_k f_{++k}\lambda_k^P \\ & - \sum_k tr\, \mathbf{W}_k^2\mathbf{X}'\mathbf{G}_k\mathbf{X} + \sum_{ijk} f_{ijk}\delta_{ij}\epsilon_i^k \\ & - \sum_{ijk} \exp\left\{\lambda + \lambda_i^R + \lambda_j^C + \lambda_k^P - tr\, \mathbf{W}_k^2\mathbf{X}'\mathbf{A}_{ij}\mathbf{X} + \delta_{ij}\epsilon_i^k\right\}, \end{aligned}$$

where $\mathbf{W}_k$ is the diagonal matrix with the weights for slice k, and $\mathbf{G}_k$ has elements g_{ijk} as defined earlier. Focusing now on the first part of the likelihood function, $\sum_{ijk} f_{ijk} \log \pi_{ijk}$ we derive the following for the coordinates

$$L = -\sum_k tr\, \mathbf{W}_k\mathbf{X}'\mathbf{G}_k\mathbf{X}\mathbf{W}_k.$$

Assuming all $\mathbf{W}_k$ are equal to the identity matrix (i.e., no time differences), and optimizing this part for $\mathbf{X}$ under $\mathbf{X}'\mathbf{X} = \mathbf{I}$, we find the maximum for $\mathbf{X}$ through an eigenvalue decomposition on the matrix $\mathbf{G}_+ = \sum_k \mathbf{G}_k$: the eigenvectors corresponding to the smallest eigenvalues give the optimum. These eigenvectors can be used as an initial estimate.

The weights: Following the same strategy as for the coordinate parameters we have to find the optimum for $-\sum_k tr\, \mathbf{W}_k^2\mathbf{A}_k$ where $\mathbf{A}_k = \mathbf{X}'\mathbf{G}_k\mathbf{X}$. If we now optimize under $1/K\sum_k \mathbf{W}_k = \mathbf{I}$, we find a solution by

$$\mathbf{W}_k = \operatorname{diag}(\mathbf{\Lambda}\mathbf{A}_k^{-1}),$$

where $\Lambda = K \times (\sum_k \text{diag}(\mathbf{A}_k)^{-1})^{-1}$.

Convergence

To check convergence we evaluate the sum of the absolute values of the likelihood equations. This sum must be smaller then a pre-specified value. The likelihood equations for the marginal parameters are given by

$$\forall i \qquad \hat{\pi}_{i++} - f_{i++} = 0,$$

$$\forall j \qquad \hat{\pi}_{+j+} - f_{+j+} = 0,$$

$$\forall k \qquad \hat{\pi}_{++k} - f_{++k} = 0,$$

For the coordinates the likelihood equation is given by

$$\sum_k \mathbf{W}_k^2 \otimes (\hat{\Gamma}_k - \mathbf{G}_k)\, \text{Vec}(\mathbf{X}) = 0,$$

where $\hat{\Gamma}_k$ has elements γ_{ijk} as defined earlier, Vec is the operation that stacks the columns of a matrix vertically, and $\otimes$ denotes the Kronecker product. For the weights the likelihood equation is given by

$$\forall k \qquad \mathbf{X}'(\hat{\Gamma}_k - \mathbf{G}_k)\mathbf{X}\mathbf{W}_k = 0.$$

After each iteration these values are checked. Futhermore, the likelihood value is checked after each iteration. If in 100 consecutive cycles the likelihood does not increase with a pre-specified small value, the algorithm is assumed to be converged.

General remarks

A small simulation study indicates that the algorithm works quite well. We made data according to a known pattern with $I = 5$ and $K = 4$. Data were produced with sample sizes of 10000, 5000, 2500, 1000, 500, and 100. With a sample size of 100 the algorithm did not work, but in that case the mean observed frequency is 1 per cell, and the number of observed zero cells is too large. In all other cases, the algorithm reproduced the generated frequencies exactly. We used one smart start and 15 random starts. In general, we think that when good starting values are provided the algorithm as described in the appendix works quite well and is stable. If starting values are bad, most often, the algorithm stops in the first few iterations because of a decrease of likelihood.

3.2 *Fit indices*

We will evaluate the models by the traditional chi-squared distributed statistics. The Pearson X^2 statistic is defined as

$$X^2 = \sum_{ijk} \frac{(f_{ijk} - \hat{\pi}_{ijk})^2}{\hat{\pi}_{ijk}},$$

where $\hat{\pi}_{ijk}$ denotes the maximum likelihood estimate of the expected frequency. The Likelihood Ratio statistic is defined as

$$LR = 2 \sum_{ijk} f_{ijk} \log \frac{f_{ijk}}{\hat{\pi}_{ijk}}.$$

The Likelihood Ratio statistic can be used to compare two models that are nested. The Likelihood Ratio statistic under the independence model gives us a measure for the total amount of association in a table. If we have a distance model in M-dimensions ($D(M)$), we can define the percentage association accounted for (%AAF, see DE ROOIJ and HEISER, 2000; GOODMAN, 1971) as

$$\%AAF = 100 \times \frac{LR_I - LR_{D(M)}}{LR_I},$$

where LR_I is the Likelihood Ratio statistic under the Independence Model, and $LR_{D(M)}$ the Likelihood Ratio statistic under a distance model in M dimensions. As is shown by De Rooij and Heiser, for large tables the chi-square distributed statistics in general dismiss a model because of the sample size. The %AAF can then be used to verify whether a model fits well enough, i.e., whether the model accounts for enough association in the data. Residuals can be used to have a closer look at the pattern of deviance, that is which cells fit the model well, and which cells do not fit the model.

4 Data analysis

As an illustration for the proposed models we will analyze the data in Table 1 with models in one- two- and three-dimensions. In our opinion, more dimensions do not give any insight into the data in a visual manner, and the advantages of our distance association models above other models disappear. First we analyzed the data using the model of independence, it strongly deviates from the data ($X^2 = 7662.15$, $LR = 5627.92$, $df = 477$). These are the base statistics against which we will compare our distance models.

We used one start with the starting values described in the previous section and 15 random starts, and selected the model with the highest likelihood from these sixteen runs. (For the three-dimensional model with diagonal parameters we had to do more random starts to obtain a solution with a higher likelihood than the three-dimensional model without diagonal parameters.) The fit statistics for models in one-, two- and three-dimensions are given in Table 2. We did analyze the data with model (1) and model (3), that is without terms for the diagonal and with one set of diagonal parameters for all K slices. The percentage association accounted for are all quite large except for the one-dimensional model without diagonal terms. The inclusion of a set of parameters for the diagonal sets increases the fit considerably. The difference

Table 2. Goodness-of-fit measures for the analysis of Table 1

Model	Measure	1 dim	2 dim	3 dim
Distance-model	X^2	3302	2122	1194
	LR	2844	1714	1192
	%AAF	49.4%	69.5%	78.8%
	df	464	451	438
Distance-model plus diagonal-term	X^2	1641	1218	1057
	LR	1500	1230	1023
	%AAF	73.3%	78.1%	81.8%
	df	454	441	428

of the two-dimensional and one-dimensional model, both with diagonal parameters, is 5% of association accounted for. Of the association not accounted for by the unidimensional model, the two-dimensional model accounts for 18%. We will discuss the solution of the two-dimensional model with one set of parameters for the diagonal in the remainder of this section.

4.1 *Interpretation of the parameters*

Here we will give a detailed discussion on the interpretation of all parameters of the selected model, the two-dimensional distance association model with one set of parameters to account for the large diagonal frequencies. We will start with the marginal parameters, then discuss the distances, the weights and the diagonal parameters. Each set of parameters pertain to a specific part of the data: the marginal parameters pertain to the main effects, the distances pertain to the association in the off-diagonal cells in every slice, and the diagonal parameters pertain to the association in the diagonal cells of every slice. We will see that all parameters have a clear and natural interpretation.

The marginal parameters

The marginal parameters provide information about the differences of occurrence of the occupational categories for the fathers (λ_i^R) and for the sons (λ_j^C), or the number of people in the different slices (λ_k^P). For example, in the first transition table a total of 15000 people were interviewed, and in the second only 7000. We do not want that difference to influence our result. This is exactly what λ_k^P accomplishes. The differences in occurrence of (occupational) categories for the fathers (or more generally, at the first time-point) over all transition tables is reflected in λ_i^R, and for the sons (or more generally, the second time-point) over all transition tables in λ_j^C. The parameters for our selected model are given in Table 3 and 4.

The marginal parameters for the fathers indicate they have the highest probability of being an unskilled or semiskilled worker (7a) and a small probability of being a lower grade technician/manual supervisor (5) or a small proprietor with employees

Table 3. Marginal and diagonal parameters for occupational categories

Parameter	1	2	3	4a	4b	5	6	7a	9c	7b
λ_i^{father}	−0.27	0.10	−0.19	−0.53	0.34	−1.11	0.57	0.72	0.34	0.02
λ_j^{son}	0.96	1.39	0.92	−0.97	−0.61	−0.56	0.99	0.88	−2.08	−0.93
ϵ_i	0.36	0.20	0.32	1.28	0.36	0.62	0.35	0.06	3.28	0.46

Table 4. Marginal parmeters for periods

	70–74	75–79	80–84	85–89	90–93
λ_k^{period}	−0.31	0.22	0.05	0.40	−0.37

(4a). For the sons, they are very unlikely to be a self-employed farmer (9c), and the highest probability is of their being a lower professional or manager (2).

The λ_k^{period} indicate the number of people asked in the different periods. These totals are 1921, 3405, 2844, 4111, and 1884 respectively. The same order as the size of the parameters.

The distances and the weights

The distances reflect the number of transitions, after taking the marginal effects out, in an inverse manner: a small distance corresponds to a large number of transitions; a large distance corresponds to a small number of transitions. The common plot with distances $d_{ij}(\mathbf{X})$ does give us exactly the information to answer the first question stated in the introduction: 'What do the patterns of intergenerational mobility look like, i.e., what is the social distance between occupational categories?'

For the data of Table 1 the distances are shown in Figure 1. We can distinguish a group consisting of large proprietors/higher professionals and managers (1), with lower professionals and managers (2) and routine non-manual workers (3) at the top indicating many transitions occur among these three categories. Another group consists of lower grade technicians/manual supervisors (5) with skilled (6) and unskilled/semi-skilled manual workers (7a). Small proprietors with (4a) and without (4b) employees are a third group. Self employed farmers together with (unskilled) agricultural workers constitute the fourth cluster. Within these clusters many transitions occur, the number of transitions between the clusters is smaller.

The vertical axis almost corresponds with the ordering as given in the table, except for the categories unskilled/semi-skilled manual worker (7a) and self employed farmers (9) which are interchanged. We can label the vertical axis as a social status dimension. The horizontal axis can be interpreted as an urbanization-axis: the desk cluster and the manual cluster against the small proprietor cluster with the agricultural cluster. The desk and manual categories are often found in the large cities; the

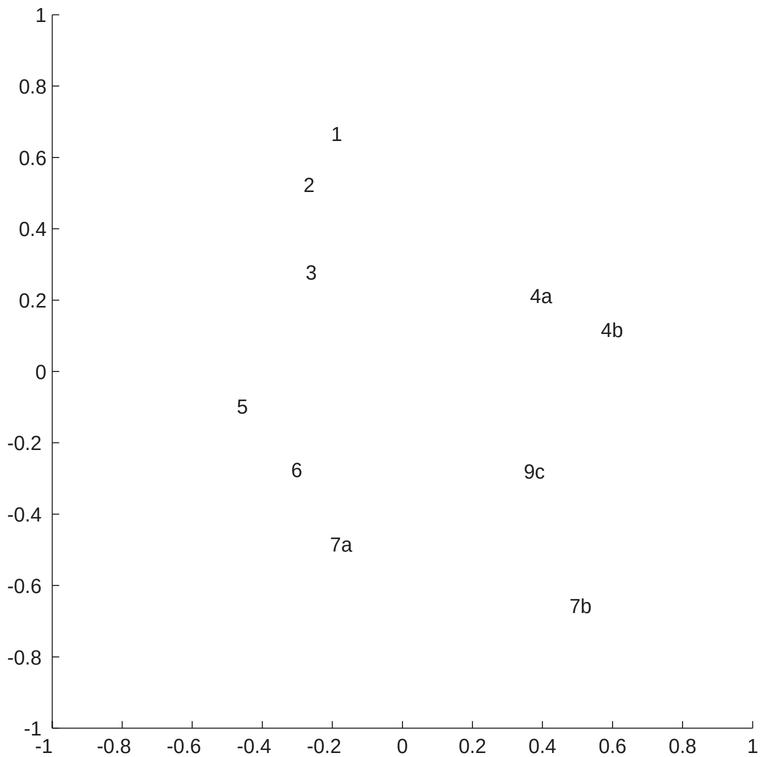

Fig. 1. The common plot for the analysis of Table 1.

farmers and small proprietors are more often found in small villages. Moreover, in the large cities, small proprietors have a difficult time because large enterprises often kill their business.

The second question: ‘Are occupational categories getting closer or further away from each other?’ is answered by the weights of each slice. If the weights become smaller the categories are getting closer together; if the weights become larger the categories are getting further apart. Figure 2 shows the weights for the analysis of Table 1 in a graphical way.

The vertical dimension becomes more important from period 1 to period 5, the weights are getting smaller, indicating the categories are getting closer. The horizontal dimension is stretched from period 1 to 5; categories are getting further apart. The vertical dimension corresponds closely to the dimension found by GANZEBOOM and LUIJKX (1995), who also found a shrinkage of this dimension. They argued that categories are getting closer together. In our result, the horizontal axis stretches. Where the groups 1, 2, 3, 4a, and 4b get closer to the other categories (5, 6, 7a, 9c, 7b), the groups 1, 2, 3, 5, 6, 7a get differentiated from the categories 4a, 4b, 9c, and 7b. The traditional ordering of the categories is vanishing, i.e., transitions between manual jobs and desk jobs occur more often. A reason might be that the manual jobs

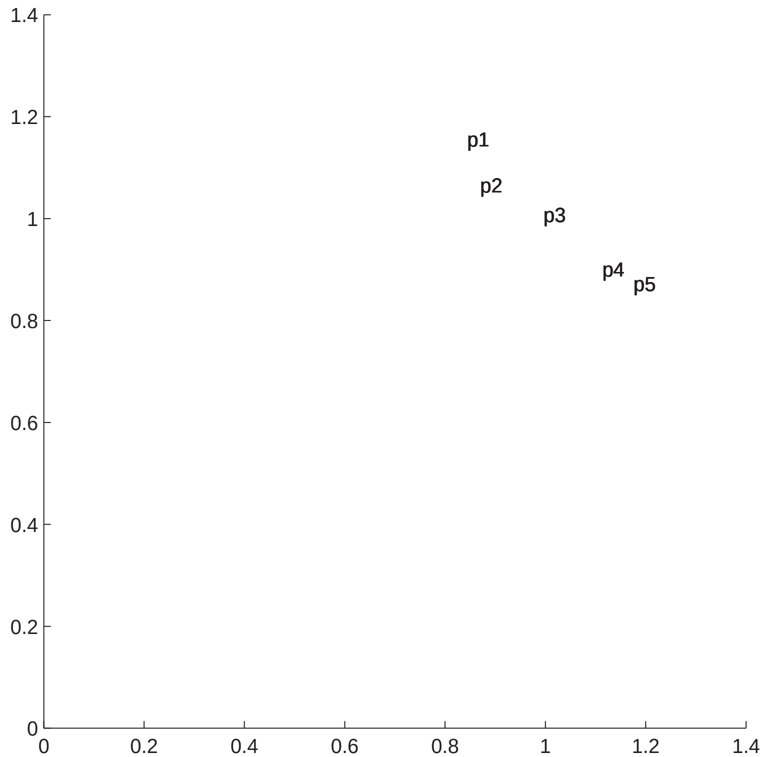

Fig. 2. The weights for the analysis of Table 1.

increasingly involve computer handling and not that much real hand craft. The horizontal dimension stretches, so the gap between people living in the city and people living in villages or in the country is widening. Over time these transitions are getting less probable.

The weights in Figure 2 are close. To check whether they have a significant contribution we performed an analysis. In this analysis we keep the weights for all slices fixed at one. As starting values for the procedure we used the results of the analysis with the distance association model in two dimensions with parameters for the diagonal (i.e., the results described in the current section). The likelihood ratio statistic for this analysis with all weights equal to one is $LR = 1324$. Comparing this likelihood ratio statistic with the one obtained earlier, we obtain $\Delta LR = 1324 - 1230 = 96$ with 8 degrees of freedom. This is a significant contribution: the weights cannot be set equal to zero without loss of information.

The diagonal terms

The diagonal parameters can be interpreted as 'immobility' parameters, or 'inheritance' parameters. A positive value means more sons follow their fathers in the corresponding occupational category than can be expected on basis of the distance part and the marginal parameters. A negative value has the reverse meaning, that is, a smaller number of sons follow their fathers than can be expected from the marginal parameters and the distance part. Since the distance between a category and itself is zero, the diagonal cells give departure from the marginal expectations, just as in the quasi-independence model. In Table 3 (last line) we give the values of the 'inheritance' parameters.

All estimates are larger than zero, indicating more sons follow their fathers than can be expected on the basis of the marginal parameters. The estimate for the self employed farmer (9c) is extremely high; many farmer's sons become farmers themselves, probably because they take over their father's farms. Few sons follow their fathers into being lower professionals or managers (2), or unskilled and semi-skilled workers (7a).

The odds ratio

In Section 2 we saw there was a direct relationship between the conditional odds ratio and the squared distances in the corresponding space (formula 6). Since the selected model includes inheritance parameters, the odds ratios have to be adapted correspondingly. As an example, let us look at the conditional odds of inheritance of being a routine non manual worker (category 3) or a small proprietor with employees (category 4a) versus changing between these two categories, i.e., the odds ratio

$$\log\left(\frac{\pi_{3,3|k} \times \pi_{4a,4a|k}}{\pi_{3,4\mathrm{a}|k} \times \pi_{4a,3|k}}\right) = 2d^2_{3,4a}(\mathbf{Y}^k) + \epsilon_3 + \epsilon_{4a}.$$

Since distances are positive by definition, and all the inheritance parameters are positive, we see that the odds are positive, and so the odds are in favor of inheritance, i.e., the probability that a son will follow his father in choice of occupation is larger than the probability that the son chooses another occupation.

The inheritage parameters are the same regardless the value of k. The conditional odds ratio changes because of the distance part. Comparing the first period with the fifth, we see that

$$d_{3,4a}(\mathbf{Y}^1) < d_{3,4a}(\mathbf{Y}^5).$$

The conditional odds ratio for the first period is smaller than the conditional odds ratio for the fifth period, indicating that a change between categories 3 and 4a is made more often in the first period than in the fifth period.

4.2 *Comparison*

In this section we will compare the results found above with results obtained using two other procedures: the weighted Euclidean model estimated by least squares methods and the multiple group association model proposed by BECKER and CLOGG (1989). In both cases we will look for a two-dimensional solution.

The weighted Euclidean model estimated by least squares

In our distance association model, the distances represent the association, so before applying a distance model we have to preprocess the data in order to delete the marginal effects. Therefore, first the observed frequency matrix is divided by the total frequency to obtain observed probabilities (p_{ijk}). From these probabilities the main effect parameters of the multiplicative model are estimated, i.e., the probabilities for the categories for the father (p_{i++}), the probabilities for the categories of the sons (p_{+j+}), and the probabilities for the different cohorts (p_{++k}), by summing. Next, we obtain residuals (θ_{ijk}) by dividing the observed probabilities by the product of the main effects

$$\theta_{ijk} = \frac{p_{ijk}}{p_{i++} \times p_{+j+} \times p_{++k}}.$$

Before applying the inverse of the Gaussian transformation to these θ_{ijk}, the θ_{ijk} should be between zero and one. Therefore, we added 0.01 to all θ_{ijk}, and then divide all by the maximum of the θ_{ijk}, i.e.,

$$\eta_{ijk} = \frac{(\theta_{ijk} + 0.01)}{\max(\theta_{ijk} + 0.01)}.$$

Then, these η_{ijk} can be transformed to dissimilarities by the inverse of the Gaussian transformation

$$\delta_{ijk} = \sqrt{(-1 \times \log(\eta_{ijk}))},$$

which can be used as dissimilarities in a multidimensional scaling procedure.

To estimate the weighted Euclidean model we used the PROXSCAL program implemented in SPSS (BUSING, COMMANDEUR, and HEISER, 1997; MEULMAN, HEISER, and SPSS INC., 1999). We used a classical scaling start, a simplex start and 50 random starts. The analysis with classical scaling starting values resulted in the final lowest stress value (normalized raw stress = 0.07807). We transformed the distances obtained through PROXSCAL back to association parameters via the Gaussian transformation. As a last step we computed parameters for the diagonal entries of the slices. Therefore, first the expected probabilities were computed by

$$\tilde{\pi}_{ijk} = p_{i++} \times p_{+j+} \times p_{++k} \times \exp(-d^2_{ijk}(\mathbf{X}; \mathbf{W}_k))$$

and then we computed p_{ii+} as follows

$$p_{ii+} = \frac{1}{K} \sum_k \left(\frac{p_{ijk}}{\tilde{\pi}_{iik}} \right).$$

Now, expected probabilities can be computed ($\hat{\pi}_{ijk}$) for model (3), and multiplied by the sample size to obtain expected frequencies under this model. The chi-square statistics are $X^2 = 2527$ and $LR = 1276$. We can compare these statistics with the statistics we found in Table 2, for the two-dimensional model plus diagonal terms. We see that the fit statistics for our distance association model in two-dimensions are better.

In Figures 3 and 4 we give the results from the PROXSCAL analysis. PROXSCAL uses different identification constraints on the weights than we do in our distance association model. We give here the PROXSCAL solution with the same constraints as in our distance association model. Comparing Figures 3 and 4 obtained here with the ones earlier (figures 1 and 2) we see that the pattern obtained with the distance association model is much more interpretable. In the PROXSCAL solution for the weights no nice trend can be seen, contrary to the weights obtained with our distance association model. The common spaces are much alike (after rotation), although our solution gives a clearer clustering onto four clusters of occupational categories as discussed in the previous section.

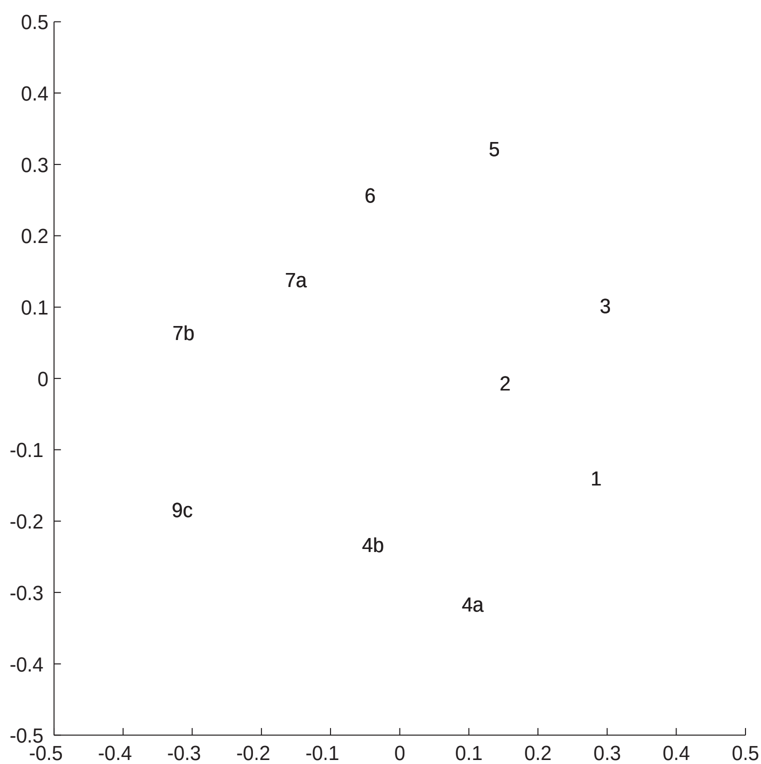

Fig. 3. The common plot obtained with PROXSCAL.

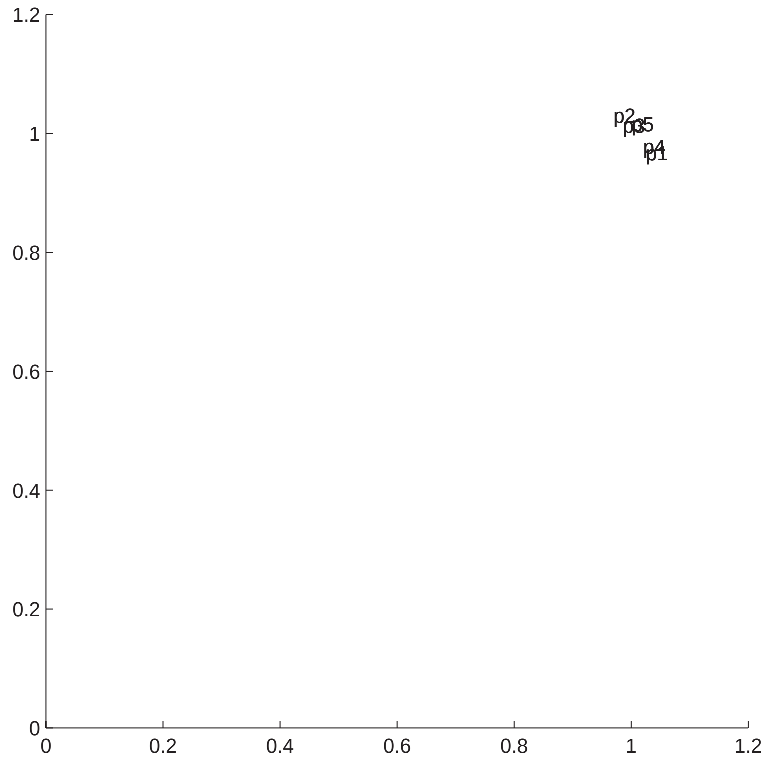

Fig. 4. The weights obtained with PROXSCAL.

The multiple group association model

In Section 2 we already discussed the multiple group association model proposed by BECKER and CLOGG (1989). Here we will compare the results obtained with that model to the results obtained with our distance association model. In our distance association model we transformed all association terms into distances. We showed that this model represents two-way association since it is a quadratic model (see Equation 4). The multiple group association model does not have such quadratic terms. Therefore the comparison of the two models cannot be totally fair. We will discuss two models, one including two-way association terms and one without. Moreover we will restrict the multiple group association model such that the row scores ($\mu_{im(k)}$) and the column scores ($\nu_{jm(k)}$) are equal for all k, just as the coordinate matrix in our distance association model is the same for every slice. The model without two-way association terms is written, in log−linear notation, as

$$\log(\pi_{ijk}) = \lambda + \lambda_i^R + \lambda_j^C + \lambda_k^P + \sum_m \phi_m^{(k)} \mu_{im} \nu_{jm} + \delta_{ij} \epsilon_i, \tag{10}$$

and the model including two-way association terms is written as

$$\log(\pi_{ijk}) = \lambda + \lambda_i^R + \lambda_j^C + \lambda_k^P + \lambda_{ik}^{RP} + \lambda_{jk}^{CP} + \sum_m \phi_m^{(k)} \mu_{im} \nu_{jm} + \delta_{ij} \epsilon_i. \quad (11)$$

In both models we can restrict the row scores (μ_{im}) to be equal to the column scores (ν_{im}) to obtain symmetric association models.

We used the program ℓEM to estimate the models (VERMUNT, 1997). In Table 5 the results of models 10 and 11 are reported. For model 10 we tried one smart start as implemented in the program and 50 random starts, in all cases the program aborted because of a decrease of the likelihood during the iterative process. Model 10 with symmetry restrictions is most comparable to our distance association model. We see that it does not fit as well as our distance association model in two dimensions. When we include association terms for the occupational categories of father and son with cohort (λ_{ik}^{RP} and λ_{jk}^{CP}, respectively), the fit gets much better, but we have to estimate much more parameters. Including these parameters in our distance association model is also possible but, as we showed in Section 2, the interpretation gets troublesome since the parameters are confounded. A word of caution is needed here on the multiple group association model without symmetry restrictions: the interpretation of the association (the relation of the row scores with the column scores) is not based on distances but on projection. We think that plots with points for both row and column categories are intuitively interpreted by distances.

5 Discussion

In the present paper we generalized the symmetric distance association model proposed by DE ROOIJ and HEISER (2000) for two-way tables, to the case of repeated transition tables. We think the model proposed is highly suitable for answering questions coming from substantive research, as illustrated in our application. Trends over time are visualized by a set of weights that easily represents changes. The common plot represents the mean transition frequency in a graphical way. The model reduces the number of parameters to estimate significantly, and allows for an easy interpretation of the final results. All parameters have a clear and straightforward interpretation.

We showed a model for the analysis of repeated transition frequency tables, where the association is transformed to a distance in Euclidean space. We transformed all two- and three-way association parameters to distances, such that the distances model

Table 5. Goodness-of-fit measures obtained with multiple group association model

Model	X^2	*LR*	*df*
model 10	**	**	**
model 10-sym	1469	1278	441
model 11	502	513	353
model 11-sym	694	659	369

the departure from independence. We could have chosen differently, for example, by also including in model (1) association parameters for the association between occupational category of the father with the period, and an association effect of occupational category for the son with the period. Such an approach will probably lead to a better model in terms of model fit. However, such an approach troubles the interpretation of the complete model, as we showed in Section 2, since those first order-interaction effect confound with the distances.

The differences between the slices k are captured in the weights. This might be considered rather restricted, because only stretching or shrinking of the dimensions is allowed. Another model, called the IDIOSCAL model (CARROLL and WISH, 1974), or the generalized Euclidean model, could also be applied . In this model, rotations of the dimensions are allowed before stretching or shrinking occurs (i.e., the matrices $\mathbf{W}_k$ are not restricted to be diagonal, but have to be nonsingular).

In the context of occupational mobility data the effects found are often symmetric (see for example, HABERMAN, 1974, Chapter 6; SOBEL, HOUT, and DUNCAN, 1985). When the association is not symmetric, a generalization we can think of is the unfolding model, in which different sets of coordinate parameters are estimated for the row categories and for the column categories. This way we allow for asymmetry in the association. The distance part of the model is then equal to

$$d^2_{ijk}(\mathbf{X};\, \mathbf{Y};\, \mathbf{W}_k) = \sum_m w^2_{km}(x_{im} - y_{jm})^2,$$

and we obtain an asymmetric association pattern since, in general, $d^2_{ijk}(\mathbf{X};\, \mathbf{Y};\, \mathbf{W}_k) \neq d^2_{ijk}(\mathbf{X};\, \mathbf{Y};\, \mathbf{W}_k)$. So, the number of people going from a to b does not need to be the same as the other way around (from b to a), even after adjustment to the marginal proportions. A model in between the Euclidean distance and the unfolding distance is the slide-vector distance, proposed by ZIELMAN and HEISER (1993). They restrict the y_{jm} to be equal to $y_{jm} = x_{jm} - \nu_m$. In this model the asymmetry is represented by a vector.

Appendix

In Section 3 we derived the first- and second-order derivatives of the log-likelihood function. Here we will show the algorithm scheme for the distance association models. In our opinion, the identification constraint on the weights (i.e., $(1/K)\sum_k \mathbf{W}_k = \mathbf{I}$) can best be implemented after each iteration. In that case the algorithm is more stable. The steps are as follows:

1. Derive starting values for all parameters, i.e., θ^0.
2. Update λ^R_i:

$$\lambda_i^{R(t+1)} = \lambda_i^{R(t)} - (\tilde{\pi}_{i++} - f_{i++})/\tilde{\pi}_{i++},$$

   where $\tilde{\pi}$ denotes the expected frequencies computed after the previous update.

3. Update λ_j^C:

$$\lambda_j^{C(t+1)} = \lambda_j^{C(t)} - (\tilde{\pi}_{+j+} - f_{+j+})/\tilde{\pi}_{+j+}.$$

4. Update λ_k^P:

$$\lambda_k^{P(t+1)} = \lambda_k^{P(t)} - (\tilde{\pi}_{++k} - f_{++k})/\tilde{\pi}_{++k}.$$

5. Update x_{im}:

$$x_{im}^{t+1} = x_{im}^t - h(x_{im})^{-1} q(x_{im}),$$

where $h(x_{im})$ is defined as in (7), and $q(x_{im})$ is defined as in (6).

6. Update w_{km}:

$$w_{km}^{t+1} = w_{km}^t - h(w_{km})^{-1} q(w_{km}),$$

where $h(w_{km})$ is defined as in (9), and $q(w_{km})$ is defined as in (8).

7. If we choose to do so, update the diagonal parameters ϵ_i^k:

$$\epsilon_i^{k(t+1)} = \epsilon_i^{k(t)} - (\tilde{\pi}_{iik} - f_{iik})/\tilde{\pi}_{iik},$$

or for ϵ_i:

$$\epsilon_i^{(t+1)} = \epsilon_i^{(t)} - (\tilde{\pi}_{ii+} - f_{ii+})/\tilde{\pi}_{ii+}.$$

8. Check convergence. If not converged go to 2, othrwise continue.
9. Compute Fit statistics.

Acknowledgement

Netherlands Organization for Scientific Research (NWO) is gratefully acknowledged for funding this project. This research was conducted while the author was supported by a grant of the Foundation for Behavioral and Educational Sciences of this organization (575-30-006).

References

AGRESTI, A. (1990), *Categorical data analysis*, John Wiley and Sons, New York.

ASHBY, F. G. (1992), *Multidimensional models of perception and recognition*, Lawrence Erlbaum Associates, Publishers, Hillsdale, New Jersey.

BECKER, M. P. (1990), Maximum likelihood estimation of the *RC(M)*-association model, *Applied Statistics* **39**, 152–167.

BECKER, M. P. and C. C. CLOGG (1989), Analysis of sets of two-way contingency tables using association models, *Journal of the American Statistical Association* **84**, 142–151.

BISHOP, Y. M. M., S. E. FIENBERG and P. W. HOLLAND (1975), *Discrete multivariate analysis: theory and practice*, MIT press, Massachusetts.

BUSING, F. M. T. A., J. J. F. COMMANDEUR and W. J. HEISER (1997), PROXSCAL: A multidimensional scaling program for individual differences scaling with constraints, in: W. BANDILLA and F. FAULBAUM (eds.), *Softstat '97: advances in statistical software 6,* Lucius & Lucius, Stuttgart, 67–74.

CARROLL, J. D. and J. J. CHANG (1970), Analysis of individual differences in multidimensional scaling via an N-way generalization of the 'Eckart–Young' decomposition, *Psychometrika* **35**, 283–319.

CARROLL, J. D. and P. E. GREEN (1997), Psychometric methods in marketing research: Part 2, multidimensional scaling, *Journal of Marketing Research* **34**, 193–204.

CARROLL, J. D. and M. WISH (1974), Models and methods for three-way multidimensional scaling, in: D. H. KRANT, R. C. ATKINSON, R. D. LUCE, and P. SUPPES (eds.), *Contemporary developments in mathematical psychology*, Freeman, San Fransisco, 57–105.

DE ROOIJ, M. (2000), *Distance models for three-way tables and three-way association: a theoretical note*, Technical Report- PRM 2000–02. Leiden University.

DE ROOIJ, M. and W. J. HEISER (2000), *Graphical representations and odds ratios in a distance association model for the analysis of cross-classified data*, Technical Report- PRM 2000–01. Leiden University.

ERIKSON, R., J. H. GOLDTHORPE and L. PORTOCARERO (1979), Intergenerational class mobility in three western European societies, *British Journal of Sociology* **30**, 415–441.

FIENBERG, S. E. (1980), *The analysis of cross-classified categorical data*, MIT Press, Massachusetts.

GANZEBOOM, H. B. G. and R. LUIJKX (1995), Intergenerationele beroepsmobiliteit in Nederland: patronen en historische veranderingen, in: J. DRONKERS and W. C. ULTEE (eds.), *Verschuivende ongelijkheid in Nederland,* Van Gorcum, Assen, 14–30.

GOODMAN, L. A. (1971), The analysis of multidimensional contingency tables: step-wise procedures and direct estimation methods for building models for multiple classifications, *Technometrics* **13**, 33–61.

GOODMAN, L. A. (1972), Some multiplicative models for the analysis of cross-classified data, *Sixth Berkeley Symposium* **1**, 649–696.

GOODMAN, L. A. (1979), Simple models for the analysis of association in cross-classifications having ordered categories, *Journal of the American Statistical Association* **74**, 537–552.

GOODMAN, L. A. (1985), The analysis of cross-classified data having ordered and/or unordered categories: association models, correlation models, and asymmetric models for contingency tables with and without missing entries, *The Annals of Statistics* **13**, 10–69.

HABERMAN, S. J. (1974), *The analysis of frequency data*, University of Chicago Press, Chicago.

KRUSKAL, J. B. (1977), Three-way arrays: rank and uniqueness of trilinear decompositions, with applications to arithmetic complexity of statistics, *Linear Algebra and its Applications* **18**, 95–138.

KRUSKAL, J. B. (1989), Rank, decompositions, and uniqueness for 3-way array and N-way array, in: R. COPPI, and S. BOLASCO, *Multiway data analysis*, North-Holland, Amsterdam, 7–18.

MCCAIN, K. W. (1998), Neural network research in context: a longitudinal journal analysis of an emerging interdisciplinary field, *Scientometrics* **41**, 389–410.

MEULMAN, J. J., W. J. HEISER, and SPSS INC. (1999), *SPSS Categories 10.0* SPSS, Chicago Illinois.

NOSOFSKY, R. (1985), Overall similarity and the identification of separable-dimension stimuli: a choice model analysis, *Perception & Psychophysics* **38**, 415–432.

RAMSAY, J. O. (1977), Maximum likelihood estimation in multidimensional scaling, *Psychometrika* **42**, 241–266.

RAMSAY, J. O. (1982), Some statistical approaches to multidimensional scaling (and discussion), *Journal of the Royal Statistical Society*, A **145**, 285–312.

RIKKEN, F., H. A. L. KIERS and R. VOS (1995), Mapping the dynamics of adverse drug reactions in subsequent time periods using INDSCAL, *Scientometrics* **33**, 367–380.

ROTH, W. M., D. GABEL, L. BROWN and D. RICE (1992), A combined method of INDSCAL and cognitive mapping for describing changes in student cognition structure, *Educational and Psychological Measurement* **52**, 769–779.

SHEPARD, R. N. (1987), Stimulus and response generalization: a stochastic model relating generalization to distance in psychological space, *Psychometrika* **22**, 325–345.

SHEPARD, R. N. (1987), Towards a universal law of generalization for psychological science, *Science* **237**, 1317–1323.

SICILIANO, R., and A. MOOIJAART (1997), Three-factor association models for three-way contingency tables, *Computational Statistics & Data Analysis* **24**, 337–356.

SOBEL, M. E., M. HOUT and O. D. DUNCAN (1985), Exchange, structure, and symmetry in occupational mobility, *American Journal of Sociology* **91**, 359–372.

TEN BERGE, J. M. F. (1991), Kruskal's polynomial for $2 \times 2 \times 2$ arrays and a generalization to $2 \times N \times N$ arrays, *Psychometrika* **56**, 631–636.

TEN BERGE, J. M. F., H. A. L. KIERS and J. DE LEEUW (1998), Explicit CANDECOMP/PARAFAC solutions for a contrived $2 \times 2 \times 2$ array of rank three, *Psychometrika* **53**, 579–584.

VERMUNT, J. K. (1997), *ℓEM*: a general program for the analysis of categorical data, The Netherlands: Tilburg University. Internet: http://www.kub.nl/faculteiten/fsw/organisatie/departementen/mto/software2.html.

WHITE, H. D. and K. W. MCCAIN (1998), Visualizing a discipline: an author cocitation analysis of information science, *Journal of the American Society for Information Science* **49**, 327–355.

ZIELMAN, B. and W. J. HEISER (1993), Analysis of asymmetry by a slide-vector, *Psychometrika* **58**, 101–114.

Received September 2000. Revised: February 2001.