RR 92-04

A LATENT CLASS UNFOLDING MODEL FOR ANALYZING SINGLE STIMULUS PREFERENCE RATINGS

Geert de Soete

University of Ghent, Belgium

Willem J. Heiser

University of Leiden, The Netherlands

The first author is supported as "Bevoegdverklaard Navorser" of the Belgian "Nationaal Fonds voor Wetenschappelijk Onderzoek".

Abstract

A multidimensional unfolding model is developed that assumes that the subjects can be clustered into a small number of homogeneous groups or classes. The subjects that belong to the same group are represented by a single ideal point. Since it is not known in advance to which group or class a subject belongs, a mixture distribution model is formulated that can be considered as a latent class model for continuous single stimulus preference ratings. A GEM algorithm is described for estimating the parameters in the model. The M–step of the algorithm is based on a majorization procedure for updating the estimates of the spatial model parameters. A strategy for selecting the appropriate number of classes and the appropriate number of dimensions is proposed and fully illustrated on some artificial data. The latent class unfolding model is applied to political science data concerning party preferences from members of the Dutch Parliament. Finally, some possible extensions of the model are discussed.

Key words: multidimensional unfolding, SMACOF, majorization method, latent class analysis, mixture distribution model, EM algorithm.

1. Introduction

Since Coombs' (1964) seminal book *A theory of data,* the unfolding model has become a popular means for constructing a geometrical representation of preference data. Let $\mathbf{Y} = ((y_{ij}))$ denote an $N \times M$ matrix containing preference ratings of N subjects for M stimuli. It will be assumed that large values of y_{ij} indicate a strong preference, while small values of y_{ij} indicate a weak preference. In a geometrical representation according to the multidimensional unfolding model, each subject i $(i = 1, \ldots, N)$ and each stimulus j $(j = 1, \ldots, M)$ is represented by a point in a joint R–dimensional space, such that the Euclidean distance between the points representing subject i and stimulus j is inversely related to y_{ij}, the preference rating of subject i for stimulus j. The points representing the subjects are often referred to as ideal points. The coordinates of the ideal point for subject i will be written as the R–component column vector $\mathbf{a}_i$, and the coordinates of the point representing stimulus j will be denoted by the R–component column vector $\mathbf{b}_j$. The $N \times R$ and $M \times R$ matrices $\mathbf{A}$ and $\mathbf{B}$ will be defined as $(\mathbf{a}_1, \ldots, \mathbf{a}_N)'$ and $(\mathbf{b}_1, \ldots, \mathbf{b}_M)'$, respectively. The Euclidean distance between $\mathbf{a}_i$ and $\mathbf{b}_j$ will be denoted by $d(\mathbf{a}_i, \mathbf{b}_j)$, and is defined as

$$d(\mathbf{a}_i, \mathbf{b}_j) = \sqrt{(\mathbf{a}_i - \mathbf{b}_j)'(\mathbf{a}_i - \mathbf{b}_j)}. \tag{1}$$

Since Kruskal's (1964) nonmetric approach to multidimensional scaling, it has become common to fit the unfolding model to a set of preference ratings $\mathbf{Y}$ by minimizing some suitably normalized version of the least squares loss function

$$L(\mathbf{A}, \mathbf{B}, \ldots) = \sum_{i=1}^{N} \sum_{j=1}^{M} [f_i(y_{ij}) - d(\mathbf{a}_i, \mathbf{b}_j)]^2 \tag{2}$$

with respect to $\mathbf{A}$, $\mathbf{B}$, and the parameters on which the transformation functions f_i depend (e.g., Kruskal, Young, & Seery, 1973). In a nonmetric unfolding method, f_i is allowed to be any monotonically decreasing function of the data of subject i. Due to the weakness of such a nonmetric model and the degeneracies that often occur when fitting this model (e.g., Heiser, 1989), attention has been mainly restricted to fitting the unfolding model to preference data that are considered to constitute an interval or a ratio scale. In the case of interval scale data with a common transformation function for all subjects, (2) reduces to

$$L(\mathbf{A}, \mathbf{B}, \alpha) = \sum_{i=1}^{N} \sum_{j=1}^{M} [\alpha - y_{ij} - d(\mathbf{a}_i, \mathbf{b}_j)]^2 , \tag{3}$$

where α denotes an additive constant. Note that the slope of the linear transformation is confounded with the scale of the configuration in the multidimensional space, and, therefore, does not need to be included explicitly in (3). In the case of ratio scale data, the additive constant α is set equal to zero. An efficient and convergent algorithm for minimizing (3) has been developed within the SMACOF framework (de Leeuw, 1977, 1988; de Leeuw & Heiser, 1977, 1980) by Heiser (1981, 1987). Note that when it is assumed that the preference ratings y_{ij} are independently normally distributed with means $\alpha - d(\mathbf{a}_i, \mathbf{b}_j)$ and a common variance, minimizing the least squares loss function (3) provides maximum likelihood estimates of the model parameters.

In this paper, we are concerned with the situation where the N subjects can be clustered into a small number of homogeneous groups such that the subjects within each group exhibit a similar preference pattern. In such a case, it is not necessary to represent each individual subject by a point in the multidimensional space. Rather, it suffices to represent the subjects within the same homogeneous group by a single ideal point. Such an approach yields a much more parsimonious representation of the structure present in the data, provided of course that the subjects do indeed cluster into a small number of homogeneous groups. In this paper, we develop a method that simultaneously arrives at a clustering of the subjects into a small number of homogeneous groups *and* constructs a geometrical representation using an unfolding model where each homogeneous group is represented by a single ideal point. The model is based on a mixture distribution formulation that can be considered as a latent class model for continuous rating data (see, De Soete, in press). Hence, the present model can be referred to as a latent class unfolding model for preference ratings. However, whereas the classical latent class model for categorical data assumes that within each latent class the data follow a product of independent Bernoulli distributions (Lazarsfeld & Henry, 1968), the present model assumes that within each homogeneous group or latent class the preference data are independently normally distributed. As remarked earlier, this assumption is consistent with the common practice of fitting the metric unfolding model by minimizing (3).

The latent class unfolding model for single stimulus preference ratings is fully stated in the next section. In Section 3, a generalized EM (or GEM) algorithm that incorporates a SMACOF method in the M–step is developed for fitting the latent class unfolding model to rating data. In Section 4, a strategy is proposed for determining the appropriate number of classes and dimensions for adequately representing a particular data set. This strategy is fully illustrated on an artificial data set. In Section 5, we present an illustrative application of the latent class unfolding model to political science data. Finally, in the last section, some possible extensions of the current approach are discussed.

2. Model

In the adaptation of the latent class model for continuous rating data (De Soete, in press), it is assumed that there exists a small number T of homogeneous groups or latent classes. Each subject i is assumed to belong to exactly one latent class t ($1 \leq t \leq T$), but it is not known in advance to which latent class a particular subject belongs. The unconditional probability that any subject i belongs to latent class t is written as λ_t, with $0 \leq \lambda_t \leq 1$ and

$$\sum_{t=1}^{T} \lambda_t = 1. \tag{4}$$

In the sequel, the T–component column vector $(\lambda_1, \ldots, \lambda_T)'$ will be referred to as $\boldsymbol{\lambda}$. It is assumed that the data of a subject i that belongs to latent class t are independently normally distributed with means $\boldsymbol{\mu}_t = (\mu_{t1}, \ldots, \mu_{tM})'$ and common variance σ^2; that is,

$$\mathbf{y}_i \sim \mathcal{N}(\boldsymbol{\mu}_t, \sigma^2 \mathbf{I}) \quad \text{for subject } i \text{ in class } t, \tag{5}$$

where $\mathbf{I}$ denotes the identity matrix and $\mathbf{y}_i$ the ith row of $\mathbf{Y}$, written as a column vector. Thus, for each latent class t, a separate mean vector $\boldsymbol{\mu}_t$ is estimated, but the variance parameter σ^2 is assumed to be common for all stimuli and all latent classes. The independence assumption entailed in (5) can be considered to be analogous to the assumption of local independence in the classical latent class model (Lazarsfeld & Henry,

1968) and in item response theory (see e.g., Lord & Novick, 1968). Note that assuming independence *within* each class does not imply that there are no correlations between the stimuli *across* the classes. Rather it implies that all the observed correlation among the stimuli can be fully explained by the fact that the subjects belong to different latent classes. Also, it should be noticed that the latent class model for continuous rating data is a special case of the latent profile model (Gibson, 1959) as operationalized by Takane (1976), and of the general mixture model of multivariate normal densities originally proposed in the psychometric literature by Wolfe (1970) and extensively discussed by McLachlan and Basford (1988).

In the latent class unfolding model the class-specific mean vector $\boldsymbol{\mu}_t$ is related to a joint geometric representation of the latent classes and the choice stimuli according to an R–dimensional Euclidean unfolding model. Let the $T \times R$ matrix $\mathbf{A} = (\mathbf{a}_1, \ldots, \mathbf{a}_T)'$ now denote the coordinates of the T latent class ideal points, and let $\mathbf{B}$ contain the stimulus coordinates as before, then the class means are assumed to be related to the Euclidean distances between the ideal points and the stimulus points as follows:

$$\mu_{tj} = \alpha_t - d(\mathbf{a}_t, \mathbf{b}_j). \tag{6}$$

As a special case, we will consider the situation where $\alpha_1 = \ldots = \alpha_T \equiv \alpha$:

$$\mu_{tj} = \alpha - d(\mathbf{a}_t, \mathbf{b}_j). \tag{7}$$

As mentioned before, response model (7) is similar to the one on which (3) is based. In many applications of multidimensional scaling it has proven useful to relate the stimulus locations to known attributes of the stimuli, or to a known design matrix according to which the stimuli were generated (see e.g., Carroll, Pruzansky, & Kruskal, 1980; De Leeuw & Heiser, 1980; Heiser & Meulman, 1983). In the present model, this amounts to requiring that the stimulus coordinates $\mathbf{B}$ are linearly related to a known $M \times G$ $(G < M)$ "design" matrix $\mathbf{X}$

$$\mathbf{B} = \mathbf{X}\mathbf{G} \tag{8}$$

where $\mathbf{G}$ is an unknown $G \times R$ transformation matrix. It can be assumed without loss of generality that $\mathbf{X}$ has full column rank.

The latent class model for rating data where the class means $\boldsymbol{\mu}_t$ are unconstrained has $T + (T \times M) + 1$ parameters (respectively for the unconditional latent class probabilities $\boldsymbol{\lambda}$, the class means $\boldsymbol{\mu}_t$, and the variance parameter σ^2). Taking constraint (4) into account, the degrees of freedom of the unconstrained latent class model for rating data are

$$T \times (M + 1). \tag{9}$$

In the latent class unfolding model, the class means $\boldsymbol{\mu}_t$ are related to a geometric representation based on the parameters $\mathbf{A}$, $\mathbf{B}$ and $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_T)'$ (in the case of (7), $\boldsymbol{\alpha}$ contains only α). Taking into account the translational and rotational indeterminacy of the Euclidean unfolding model, the degrees of freedom of the latent class unfolding model become

$$2T + (T + M)R - \frac{R(R+1)}{2} \tag{10}$$

when a separate additive constant α_t is estimated per latent class, and

$$T + 1 + (T + M)R - \frac{R(R+1)}{2} \tag{11}$$

when a common additive constant is estimated for all classes. When linear constraints are imposed on the stimulus coordinates through (8), $\mathbf{G}$ needs to be estimated instead of $\mathbf{B}$. Also note that imposing linear constraints on $\mathbf{B}$ removes the translational indeterminacy in the Euclidean unfolding model. Hence, in the linearly constrained case combined with (6), the model degrees of freedom are

$$2T + (T + G)R - \frac{R(R-1)}{2}, \tag{12}$$

while the degrees of freedom of the linearly constrained latent class unfolding model based on (7) are

$$T + 1 + (T + G)R - \frac{R(R-1)}{2}. \tag{13}$$

Of course, the latent class unfolding model only imposes effective constraints on the latent class means if its degrees of freedom are smaller than (9). This imposes an upper limit on the number of dimensions R in the model.

3. Parameter Estimation

The Likelihood Function

Equation (5) implies that the probability density function (p.d.f.) of the data, $\mathbf{y}_i$, of a subject i that is known to belong to latent class t is

$$f(\mathbf{y}_i|\mathbf{a}_t,\mathbf{B},\alpha_t,\sigma^2) = (\sigma\sqrt{2\pi})^{-M}\exp\left[-\frac{(\mathbf{y}_i-\boldsymbol{\mu}_t)'(\mathbf{y}_i-\boldsymbol{\mu}_t)}{2\sigma^2}\right] \tag{14}$$

(with $\mathbf{G}$ replacing $\mathbf{B}$ in the case where linear constraints are imposed on $\mathbf{B}$). However, in practice we do not know in advance to which latent class a particular subject i belongs. Hence, the p.d.f. of the data of an arbitrary subject i becomes a finite mixture of multivariate normal densities of the form (14):

$$\begin{aligned} g(\mathbf{y}_i|\mathbf{A},\mathbf{B},\boldsymbol{\alpha},\sigma^2,\boldsymbol{\lambda}) &= \sum_{t=1}^{T}\lambda_t f(\mathbf{y}_i|\mathbf{a}_t,\mathbf{B},\alpha_t,\sigma^2) \\ &= (\sigma\sqrt{2\pi})^{-M}\sum_{t=1}^{T}\lambda_t\exp\left[-\frac{(\mathbf{y}_i-\boldsymbol{\mu}_t)'(\mathbf{y}_i-\boldsymbol{\mu}_t)}{2\sigma^2}\right]. \end{aligned} \tag{15}$$

Maximum likelihood estimates of the model parameters $\mathbf{A}$, $\mathbf{B}$ (or $\mathbf{G}$ in the linearly constrained case), $\boldsymbol{\alpha}$, σ^2, and $\boldsymbol{\lambda}$, can be obtained by maximizing the likelihood function

$$\begin{aligned} L(\mathbf{A},\mathbf{B},\boldsymbol{\alpha},\sigma^2,\boldsymbol{\lambda}|\mathbf{Y}) &= \prod_{i=1}^{N} g(\mathbf{y}_i|\mathbf{A},\mathbf{B},\boldsymbol{\alpha},\sigma^2,\boldsymbol{\lambda}) \\ &= (\sigma\sqrt{2\pi})^{-N\cdot M}\prod_{i=1}^{N}\left\{\sum_{t=1}^{T}\lambda_t\exp\left[-\frac{(\mathbf{y}_i-\boldsymbol{\mu}_t)'(\mathbf{y}_i-\boldsymbol{\mu}_t)}{2\sigma^2}\right]\right\} \end{aligned} \tag{16}$$

subject to (4). Once maximum likelihood estimates $\hat{\mathbf{A}}$, $\hat{\mathbf{B}}$, $\hat{\boldsymbol{\alpha}}$, $\hat{\sigma}^2$ and $\hat{\boldsymbol{\lambda}}$ are available, the a posteriori probability that a subject i belongs to latent class t can be computed by means of Bayes' theorem. This a posteriori probability, written as $h_{it}(\hat{\mathbf{A}},\hat{\mathbf{B}},\hat{\boldsymbol{\alpha}},\hat{\sigma}^2,\hat{\boldsymbol{\lambda}})$, equals

$$h_{it}(\hat{\mathbf{A}},\hat{\mathbf{B}},\hat{\boldsymbol{\alpha}},\hat{\sigma}^2,\hat{\boldsymbol{\lambda}}) = \frac{\hat{\lambda}_t f(\mathbf{y}_i|\hat{\mathbf{a}}_t,\hat{\mathbf{B}},\hat{\alpha}_t,\hat{\sigma}^2)}{g(\mathbf{y}_i|\hat{\mathbf{A}},\hat{\mathbf{B}},\hat{\boldsymbol{\alpha}},\hat{\sigma}^2,\hat{\lambda})}$$

$$= \frac{\hat{\lambda}_t \exp\left[-\frac{(\mathbf{y}_i - \hat{\boldsymbol{\mu}}_t)'(\mathbf{y}_i - \hat{\boldsymbol{\mu}}_t)}{2\hat{\sigma}^2}\right]}{\sum_{s=1}^{T} \hat{\lambda}_s \exp\left[-\frac{(\mathbf{y}_i - \hat{\boldsymbol{\mu}}_s)'(\mathbf{y}_i - \hat{\boldsymbol{\mu}}_s)}{2\hat{\sigma}^2}\right]}, \tag{17}$$

where the jth component of $\hat{\boldsymbol{\mu}}_t$ is defined as

$$\hat{\mu}_{tj} = \hat{\alpha}_t - d(\hat{\mathbf{a}}_t, \hat{\mathbf{b}}_j). \tag{18}$$

A subject can then be assigned to the class he or she is most likely to belong to, given these posterior probabilities. A classification based on such a posteriori class membership probabilities is carried out in the application reported in Section 5.

GEM Algorithm

As is the case with most mixture distribution problems (see, McLachlan & Basford 1988), the parameters of the latent class unfolding model are most easily estimated by means of an EM algorithm (Dempster, Laird, & Rubin, 1977). In order to enable an EM algorithm formulation, the following non-observed data are introduced:

$$z_{it} = \begin{cases} 1 & \textit{iff} \text{ subject } i \text{ belongs to latent class } t, \\ 0 & \text{otherwise.} \end{cases} \tag{19}$$

It will be convenient to define the column vector $\mathbf{z}_i$ as $(z_{i1}, \ldots, z_{iT})'$ and the N by T matrix as $\mathbf{Z}$ as $(\mathbf{z}_1, \ldots, \mathbf{z}_N)'$. It is assumed that the non-observed data $\mathbf{z}_i$ are independently and identically multinomially distributed with probabilities $\boldsymbol{\lambda}$; that is,

$$(\mathbf{z}_i|\boldsymbol{\lambda}) \sim \prod_{t=1}^{T} \lambda_t^{z_{it}}. \tag{20}$$

The distribution of $\mathbf{y}_i$, given $\mathbf{z}_i$, can be written as

$$(\mathbf{y}_i|\mathbf{z}_i, \mathbf{A}, \mathbf{B}, \boldsymbol{\alpha}, \sigma^2, \boldsymbol{\lambda}) \sim \sum_{t=1}^{T} z_{it} f(\mathbf{y}_i|\mathbf{a}_t, \mathbf{B}, \alpha_t, \sigma^2),$$

which is equivalent to

$$(\mathbf{y}_i|\mathbf{z}_i, \mathbf{A}, \mathbf{B}, \boldsymbol{\alpha}, \sigma^2, \boldsymbol{\lambda}) \sim \prod_{t=1}^{T} f(\mathbf{y}_i|\mathbf{a}_t, \mathbf{B}, \alpha_t, \sigma^2)^{z_{it}}. \tag{21}$$

It follows from (20) and (21) that the log-likelihood of the complete data $\mathbf{Y}$ and $\mathbf{Z}$ can be written as

$$\log L_C(\mathbf{A}, \mathbf{B}, \boldsymbol{\alpha}, \sigma^2, \boldsymbol{\lambda}|\mathbf{Y}, \mathbf{Z}) = \sum_{i=1}^{N}\sum_{t=1}^{T} z_{it} \log f(\mathbf{y}_i|\mathbf{a}_t, \mathbf{B}, \alpha_t, \sigma^2) + \sum_{i=1}^{N}\sum_{t=1}^{T} z_{it} \log \lambda_t. \tag{22}$$

An EM algorithm for maximizing (16) alternates iteratively between an E–step (expectation step) and an M–step (maximization step). In the E–step, the expectation of $\log L_C$ needs to be calculated over the conditional distribution of the non-observed data $\mathbf{Z}$, given the observed data $\mathbf{Y}$ and provisional estimates $\mathbf{A}^{(0)}, \mathbf{B}^{(0)}, \boldsymbol{\alpha}^{(0)}, \sigma^{2(0)}, \boldsymbol{\lambda}^{(0)}$ of the model parameters. This expectation is

$$\begin{aligned} Q(\mathbf{A}, \mathbf{B}, \boldsymbol{\alpha}, \sigma^2, \boldsymbol{\lambda}, \mathbf{A}^{(0)}, \mathbf{B}^{(0)}, \boldsymbol{\alpha}^{(0)}, \sigma^{2(0)}, \boldsymbol{\lambda}^{(0)}) = \\ \sum_{i=1}^{N}\sum_{t=1}^{T} E(z_{ij}|\mathbf{A}^{(0)}, \mathbf{B}^{(0)}, \boldsymbol{\alpha}^{(0)}, \sigma^{2(0)}, \boldsymbol{\lambda}^{(0)}) \log f(\mathbf{y}_i|\mathbf{a}_t, \mathbf{B}, \alpha_t, \sigma^2) \\ + \sum_{i=1}^{N}\sum_{t=1}^{T} E(z_{it}|\mathbf{A}^{(0)}, \mathbf{B}^{(0)}, \boldsymbol{\alpha}^{(0)}, \sigma^{2(0)}, \boldsymbol{\lambda}^{(0)}) \log \lambda_t. \end{aligned} \tag{23}$$

In the M–step, $Q(\mathbf{A}, \mathbf{B}, \boldsymbol{\alpha}, \sigma^2, \boldsymbol{\lambda}, \mathbf{A}^{(0)}, \mathbf{B}^{(0)}, \boldsymbol{\alpha}^{(0)}, \sigma^{2(0)}, \boldsymbol{\lambda}^{(0)})$ must be maximized with respect to $\mathbf{A}, \mathbf{B}, \boldsymbol{\alpha}, \sigma^2, \boldsymbol{\lambda}$ to obtain new provisional parameter estimates. Whereas in the M–step of a regular EM algorithm, (23) is completely maximized with respect to the model parameters, in a generalized EM (or GEM) algorithm, the M–step determines new estimates of the model parameters that improve upon the previous estimates $\mathbf{A}^{(0)}, \mathbf{B}^{(0)}, \boldsymbol{\alpha}^{(0)}, \sigma^{2(0)}, \boldsymbol{\lambda}^{(0)}$ without necessarily minimizing (23) completely (see, Dempster et al., 1977).

E–Step

The expectation of the non-observed data given provisional estimates of the model parameters can be readily determined

$$E(z_{it}|\mathbf{A}^{(0)},\mathbf{B}^{(0)},\boldsymbol{\alpha}^{(0)},\sigma^{2(0)},\boldsymbol{\lambda}^{(0)}) = h_{it}(\mathbf{A}^{(0)},\mathbf{B}^{(0)},\boldsymbol{\alpha}^{(0)},\sigma^{2(0)},\boldsymbol{\lambda}^{(0)}). \tag{24}$$

Thus, in the E–step, the non-observed data $\mathbf{Z}$ are replaced by the posterior probabilities, defined in (17), calculcated on the basis of the provisional parameter estimates $\mathbf{A}^{(0)},\mathbf{B}^{(0)},\boldsymbol{\alpha}^{(0)},\sigma^{2(0)},\boldsymbol{\lambda}^{(0)}$. Denoting (24) by $z_{it}^{(0)}$, (23) can be written as

$$\begin{aligned} Q(\mathbf{A},\mathbf{B},\boldsymbol{\alpha},\sigma^2,\boldsymbol{\lambda},\mathbf{A}^{(0)},\mathbf{B}^{(0)},\boldsymbol{\alpha}^{(0)},\sigma^{2(0)},\boldsymbol{\lambda}^{(0)}) &= \sum_{i=1}^{N}\sum_{t=1}^{T} z_{it}^{(0)} \log f(\mathbf{y}_i|\mathbf{a}_t,\mathbf{B},\alpha_t,\sigma^2) \\ &\quad + \sum_{i=1}^{N}\sum_{t=1}^{T} z_{it}^{(0)} \log \lambda_t. \end{aligned} \tag{25}$$

M–Step

It can be easily shown that, in order to maximize (25) with respect to $\mathbf{A}$, $\mathbf{B}$, and $\boldsymbol{\alpha}$, it suffices to minimize

$$\begin{aligned} q(\mathbf{A},\mathbf{B},\boldsymbol{\alpha}) &= \sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{j=1}^{M} z_{it}^{(0)}(y_{ij}-\mu_{tj})^2 \\ &= \sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{j=1}^{M} z_{it}^{(0)}(y_{ij}-\bar{y}_{tj})^2 + \phi(\mathbf{A},\mathbf{B},\boldsymbol{\alpha}), \end{aligned} \tag{26}$$

where $\bar{y}_{tj}$ is defined as

$$\bar{y}_{tj} = \frac{\sum_{i=1}^{N} z_{it}^{(0)} y_{ij}}{\sum_{i=1}^{N} z_{it}^{(0)}}, \tag{27}$$

and the function $\phi(\mathbf{A},\mathbf{B},\boldsymbol{\alpha})$ as

$$\phi(\mathbf{A},\mathbf{B},\boldsymbol{\alpha}) = \sum_{t=1}^{T}\sum_{j=1}^{M} \gamma_t(\bar{y}_{tj}-\mu_{tj})^2 \tag{28}$$

with

$$\gamma_t = \sum_{i=1}^{N} z_{it}^{(0)}. \tag{29}$$

As in (26) $q(\mathbf{A}, \mathbf{B}, \boldsymbol{\alpha})$ is orthogonally decomposed into a within-class and a between-class component, it is minimized with respect to $\mathbf{A}$, $\mathbf{B}$, and $\boldsymbol{\alpha}$, by minimizing $\phi(\mathbf{A}, \mathbf{B}, \boldsymbol{\alpha})$. Defining the pseudodistances $\tilde{d}_{tj}$ as

$$\tilde{d}_{tj} = \alpha_t - \bar{y}_{tj} \tag{30}$$

(or, as $\alpha - \bar{y}_{tj}$ in the case of (7)), the weighted least squares function $\phi(\mathbf{A}, \mathbf{B}, \boldsymbol{\alpha})$ can be written as

$$\phi(\mathbf{A}, \mathbf{B}, \boldsymbol{\alpha}) = \sum_{t=1}^{T} \sum_{j=1}^{M} \gamma_t [\tilde{d}_{tj} - d(\mathbf{a}_t, \mathbf{b}_j)]^2. \tag{31}$$

From (31) it is obvious that ϕ can be minimized by means of the SMACOF algorithm originally proposed by de Leeuw and Heiser (de Leeuw, 1977, 1988; de Leeuw & Heiser, 1977, 1980) and adapted to the unfolding case by Heiser (1981, 1987). However, since the pseudo-distances $\tilde{d}_{tj}$ may be negative, it is necessary to take into account the modifications of the SMACOF algorithm that were recently suggested by Heiser (1991a). A SMACOF algorithm for minimizing ϕ will be outlined here without proof. For a statement of the relevant proofs, the reader is referred to de Leeuw (1988), de Leeuw and Heiser (1980) and Heiser (1991a).

Given a current estimates $\mathbf{A}^{(0)}$, $\mathbf{B}^{(0)}$ and $\boldsymbol{\alpha}^{(0)}$ of the coordinates and the additive constants, we define the $T \times M$ matrices $\mathbf{W} = ((w_{tj}))$ and $\mathbf{C} = ((c_{tj}))$ as follows (Heiser, 1991a, p. 18):

$$w_{tj} = \begin{cases} \gamma_t & \textit{iff } d(\mathbf{a}_t{}^{(0)}, \mathbf{b}_j{}^{(0)}) > 0 \text{ and } \tilde{d}_{tj}^{(0)} \geq 0, \\ \dfrac{\gamma_t[d(\mathbf{a}_t{}^{(0)}, \mathbf{b}_j{}^{(0)}) + |\tilde{d}_{tj}^{(0)}|]}{d(\mathbf{a}_t{}^{(0)}, \mathbf{b}_j{}^{(0)})} & \textit{iff } d(\mathbf{a}_t{}^{(0)}, \mathbf{b}_j{}^{(0)}) > 0 \text{ and } \tilde{d}_{tj}^{(0)} < 0, \\ \dfrac{\gamma_t(\epsilon + |\tilde{d}_{tj}^{(0)}|^2)}{\epsilon} & \textit{iff } d(\mathbf{a}_t{}^{(0)}, \mathbf{b}_j{}^{(0)}) = 0 \text{ and } \tilde{d}_{tj}^{(0)} < 0, \end{cases} \tag{32}$$

and

$$c_{tj} = \begin{cases} \gamma_t \tilde{d}_{tj}^{(0)} / d(\mathbf{a}_t{}^{(0)}, \mathbf{b}_j{}^{(0)}) & \textit{iff } d(\mathbf{a}_t{}^{(0)}, \mathbf{b}_j{}^{(0)}) > 0 \text{ and } \tilde{d}_{tj}^{(0)} \geq 0, \\ 0 & \text{otherwise,} \end{cases} \tag{33}$$

where ϵ is a small positive constant and where $\tilde{d}_{tj}^{(0)}$ is defined by

$$\tilde{d}_{tj}^{(0)} = \alpha_t^{(0)} - \bar{y}_{tj}. \tag{34}$$

We also need the $T \times T$ diagonal matrices $\mathbf{P}$ and $\mathbf{R}$ containing on the main diagonal the row sums of $\mathbf{C}$ and $\mathbf{W}$, respectively:

$$\mathbf{P} = \text{diag}(\mathbf{C}\mathbf{1}_M) \tag{35}$$

and

$$\mathbf{R} = \text{diag}(\mathbf{W}\mathbf{1}_M) \tag{36}$$

where the notation $\mathbf{1}_n$ is used to denote an n–component column vector of ones. Similarly, we introduce the $M \times M$ diagonal matrices $\mathbf{E}$ and $\mathbf{F}$ containing the column sums of $\mathbf{C}$ and $\mathbf{W}$:

$$\mathbf{E} = \text{diag}(\mathbf{1}_T'\mathbf{C}) \tag{37}$$

and

$$\mathbf{F} = \text{diag}(\mathbf{1}_T'\mathbf{W}). \tag{38}$$

It can be shown (Heiser, 1981, 1987) that new estimates of $\mathbf{A}$ and $\mathbf{B}$ that improve upon the previous estimates $\mathbf{A}^{(0)}$ and $\mathbf{B}^{(0)}$ are the solution of the following system of linear equations:

$$\begin{aligned} \mathbf{RA} - \mathbf{WB} &= \tilde{\mathbf{A}} \\ \mathbf{FB} - \mathbf{W}'\mathbf{A} &= \tilde{\mathbf{B}}, \end{aligned} \tag{39}$$

with $\tilde{\mathbf{A}}$ and $\tilde{\mathbf{B}}$ defined by

$$\begin{aligned} \tilde{\mathbf{A}} &= \mathbf{PA}^{(0)} - \mathbf{CB}^{(0)} \\ \tilde{\mathbf{B}} &= \mathbf{EB}^{(0)} - \mathbf{C}'\mathbf{A}^{(0)}. \end{aligned} \tag{40}$$

Since in most applications T is smaller than M, (39) is most efficiently solved by first solving

$$(\mathbf{R} - \mathbf{WF}^{-1}\mathbf{W}')\mathbf{A} = \tilde{\mathbf{A}} + \mathbf{WF}^{-1}\tilde{\mathbf{B}} \tag{41}$$

with respect to $\mathbf{A}$, and then determining a new estimate $\hat{\mathbf{B}}$ of $\mathbf{B}$ from

$$\hat{\mathbf{B}} = \mathbf{F}^{-1}(\tilde{\mathbf{B}} + \mathbf{W}'\hat{\mathbf{A}}), \tag{42}$$

where $\hat{\mathbf{A}}$ denotes the solution of (41). When linear constraints are imposed on the stimulus coordinates, the approach described above cannot be applied. In the linearly constrained case, a new estimate of $\mathbf{G}$ is obtained my minimizing

$$\text{tr}\left[\mathbf{XG} - \mathbf{F}^{-1}(\tilde{\mathbf{B}} + \mathbf{W}'\mathbf{A}^{(0)})\right]'\mathbf{F}\left[(\mathbf{XG} - \mathbf{F}^{-1}(\tilde{\mathbf{B}} + \mathbf{W}'\mathbf{A}^{(0)})\right] \tag{43}$$

with respect to $\mathbf{G}$ (see, Heiser, 1987, Section 3.6). The solution of this weighted regression problem is well-known:

$$\hat{\mathbf{G}} = (\mathbf{X}'\mathbf{FX})^{-1}\mathbf{X}'(\tilde{\mathbf{B}} + \mathbf{W}'\mathbf{A}^{(0)}). \tag{44}$$

A new estimate of $\mathbf{A}$ can then be obtained from

$$\hat{\mathbf{A}} = \mathbf{R}^{-1}(\tilde{\mathbf{A}} + \mathbf{WX}\hat{\mathbf{G}}). \tag{45}$$

It can be shown that the new estimates of $\mathbf{A}$ and $\mathbf{B}$ (or $\mathbf{G}$) always satisfy

$$\phi(\hat{\mathbf{A}}, \hat{\mathbf{B}}, \boldsymbol{\alpha}^{(0)}) \leq \phi(\mathbf{A}^{(0)}, \mathbf{B}^{(0)}, \boldsymbol{\alpha}^{(0)}),$$

where equality holds only if $\hat{\mathbf{A}} = \mathbf{A}^{(0)}$ and $\hat{\mathbf{B}} = \mathbf{B}^{(0)}$ (see, de Leeuw & Heiser, 1980; Heiser, 1991a).

Given estimates of the ideal point and stimulus point coordinates, new estimates of the additive constants α_t can be obtained through

$$\hat{\alpha}_t = \frac{1}{M}\sum_{j=1}^{M}[\bar{y}_{tj} + d(\hat{\mathbf{a}}_t, \hat{\mathbf{b}}_j)]. \tag{46}$$

In the case of a common additive constant for all T latent classes, a new estimate of the common additive constant is provided by

$$\hat{\alpha} = \frac{\sum_{t=1}^{T}\sum_{j=1}^{M} \gamma_t[\bar{y}_{tj} + d(\hat{\mathbf{a}}_t, \hat{\mathbf{b}}_j)]}{M \sum_{t=1}^{T} \gamma_t}. \tag{47}$$

A regular SMACOF algorithm for minimizing (31) would alternate between obtaining new estimates of $\mathbf{A}$ and $\mathbf{B}$ and new estimates of α_t in an iterative fashion until no further improvement is possible. In an M–step of a GEM algorithm it suffices to obtain estimates of $\mathbf{A}$, $\mathbf{B}$ and $\boldsymbol{\alpha}$ that improve upon the previous estimates. In practice, we carry out a few alternating (weighted) least squares cycles that alternate between updating $\mathbf{A}$ and $\mathbf{B}$ and updating $\boldsymbol{\alpha}$.

In the M–step we also need to obtain new estimates of σ^2 and $\boldsymbol{\lambda}$. Setting the partial derivative of $Q(\mathbf{A}, \mathbf{B}, \boldsymbol{\alpha}, \sigma^2, \boldsymbol{\lambda}, \mathbf{A}^{(0)}, \mathbf{B}^{(0)}, \boldsymbol{\alpha}^{(0)}, \sigma^{2(0)}, \boldsymbol{\lambda}^{(0)})$ with respect to σ^2 equal to zero yields

$$\hat{\sigma}^2 = \frac{1}{NM} \sum_{i=1}^{N}\sum_{t=1}^{T} z_{it}^{(0)} (\mathbf{y}_i - \bar{\mathbf{y}}_t)'(\mathbf{y}_i - \bar{\mathbf{y}}_t) \tag{48}$$

with $\bar{\mathbf{y}}_t = (\bar{y}_{t1}, \ldots, \bar{y}_{tM})'$. Maximizing $Q(\mathbf{A}, \mathbf{B}, \boldsymbol{\alpha}, \sigma^2, \boldsymbol{\lambda}, \mathbf{A}^{(0)}, \mathbf{B}^{(0)}, \boldsymbol{\alpha}^{(0)}, \sigma^{2(0)}, \boldsymbol{\lambda}^{(0)})$ with respect to $\boldsymbol{\lambda}$ subject to (4) gives

$$\hat{\lambda}_t = \frac{1}{N} \sum_{i=1}^{N} z_{it}^{(0)}. \tag{49}$$

Initial Parameter Estimates

The GEM algorithm for estimating the parameters in the latent class unfolding model alternatives between applying an E–step and an M–step in an iterative fashion, starting from some initial parameter estimates. The iterative process is continued until no further improvement in the likelihood function (16) is possible. Initial estimates of the non-observed data $\mathbf{Z}$ are obtained from a k–means clustering on $\mathbf{Y}$. From the cluster means for the M stimuli, initial estimates are derived of $\mathbf{A}$ and $\mathbf{B}$ using the heuristic procedure described in Section 2.4 of Heiser (1981). An initial estimate of $\boldsymbol{\alpha}$ is chosen such that all $\tilde{d}_{tj}$ are nonnegative.

4. Model Selection

When applying the latent class unfolding model, the number of classes, T, and the number of dimensions, R, need to be specified. In this section, a strategy from selecting the appropriate values of T and R is described.

We propose to determine the appropriate number of classes from comparisons of the goodness of fit of the latent class model (for rating data) with no constraints on the class means $\boldsymbol{\mu}_t$, for a varying number of latent classes T. The usual procedure for comparing the goodness of fit of a T-class unconstrained model versus a $(T+1)$-class unconstrained model involves assessing the significance of the likelihood ratio statistic

$$U = -2\log\left(\frac{\hat{L}^{(T)}}{\hat{L}^{(T+1)}}\right) \tag{50}$$

where $\hat{L}^{(T)}$ and $\hat{L}^{(T+1)}$ denote the maximum of the likelihood function obtained for the T-class and $(T+1)$-class model, respectively. Under the null hypothesis that a T-class model fits the data equally well as a $(T+1)$-class model, U would be asymptotically distributed as a chi-square with degrees of freedom equal to the difference in degrees of freedom on the two models, *if* certain regularity conditions were satisfied. Unfortunately, when comparing two mixture models with a different number of component distributions, the regularity conditions are known to be violated (see, McLachlan & Basford, 1988), and the distribution of the likelihood ratio statistic U is not known. As an alternative to relying on asymptotic maximum likelihood theory, the sampling distribution of U can be determined by means of a Monte Carlo procedure originally suggested by Hope (1968) and successfully applied in various latent class applications (e.g., De Soete, 1990; De Soete & DeSarbo, 1991; De Soete & Winsberg, in press).

Let $\hat{\boldsymbol{\lambda}}$, $\hat{\boldsymbol{\mu}}_1, \ldots, \hat{\boldsymbol{\mu}}_T$ and $\hat{\sigma}^2$ denote the maximum likelihood estimates of the parameters in the T-class model, derived from the data $\mathbf{Y}$. From the T-class population with parameters $\hat{\boldsymbol{\lambda}}$, $\hat{\boldsymbol{\mu}}_1, \ldots, \hat{\boldsymbol{\mu}}_T$ and $\hat{\sigma}^2$, $n-1$ random samples $\mathbf{Y}^*$ of size N are drawn. Both the T-class and the $(T+1)$-class unconstrained models are fit to each Monte Carlo sample $\mathbf{Y}^*$, and the value of the likelihood ratio statistic U is computed. The null hypothesis that the T-class model fits the data equally well as the $(T+1)$-class model is rejected at significance level p, whenever the value of U for the observed data $\mathbf{Y}$ exceeds

$n(1-p)$ of the values of U obtained for the Monte Carlo samples $\mathbf{Y}^*$. A minimal value of n when adopting a significance level $p = 0.05$ is 20. Hope (1968) proved that the power of this Monte Carlo significance test increases as n increases. As McLachlan (1987) argued, this procedure can be considered as a *parametric* bootstrap method.

Once the appropriate number of classes, $\hat{T}$, has been determined by means of the Monte Carlo procedure outlined above, the goodness of fit of the $\hat{T}$–class unfolding model with varying dimensionality R can be compared with the goodness of fit of the $\hat{T}$–class latent class model where no constraints are imposed on the class means $\boldsymbol{\mu}_t$, to determine the appropriate number of dimensions. Since the models being compared have the same number of latent classes, the regularity conditions are satisfied and the asymptotic distribution of the relevant likelihood ratio statistic is known. However, experience with various artificial data sets revealed that, in some cases, the distribution of U does not approximate its asymptotic distribution well, unless the number of subjects, N, is fairly large. Therefore, when N is not very large, we suggest to rely for these comparisons on the Monte Carlo significance procedure described earlier.

Insert Table 1 about here

For illustrative purposes, we apply this model selection strategy to some artificial data. To generate the data, sixteen stimulus points were located according to a 4×4 equally spaced grid in a two-dimensional space. Three latent class ideal points were located in the interior of the grid. The distances computed from this configuration were converted into latent class means according to (7). The variance parameter σ^2 was chosen such that about 25% of the total variance of the generated data was error variance. Fifty subjects were sampled from each latent class population, resulting in a 150×16 data matrix. To determine the appropriate number of latent classes for this data set, the latent class model with no constraints on the class means was fit with T ranging from 1 to 4. The improvement of a $(T+1)$–class solution over a T–class solution was assessed by means of the Monte Carlo significance procedure with $n = 500$. The results of these analyses are reported in the upper part of Table 1. It can be inferred from the Monte Carlo significance tests that a 3–class unconstrained model fits the data equally well as 4–class model, while a 3–class model fits the data better than a 2–class model. Hence, the appropriate number of classes appears to be three, which is equal to the true number of classes from which the data were generated. To illustrate that when a 3–class unconstrained model is compared with a 4–class unconstrained model, the likelihood ratio statistic is not distributed as a chi square with 15 degrees of freedom (which is the difference in degrees of freedom of the 4–class and the 3–class model), we estimated the sampling distribution of U from the values of U obtained for the Monte Carlo samples, using a nonparametric density estimation method (Wegman, 1972). The resulting estimate is presented in Figure 1 along with a chi square distribution with 15 degrees of freedom. From the figure it is obvious that, in this case, the likelihood ratio statistic is not distributed as a chi square statistic with 15 degrees of freedom.

Insert Figure 1 about here

Subsequently, the 3–class two-dimensional unfolding model with (7) was fit to the data and compared with the 3–class class model with no constraints on the latent class means. This model comparison is reported in the lower part of Table 1. Since the likelihood ratio statistic is not significant, it can be concluded that a 3–class two-dimensional unfolding model fits the data equally well as a 3–class unconstrained model. Finally, we fitted the 3–class two-dimensional unfolding model with linear constraints on the stimulus coordinates as specified in (8). For $\mathbf{X}$, we used a design matrix that encodes the four by four design from which the stimulus points were generated. This design matrix is presented in Table 2. This constrained two-dimensional model was compared with the 3–class model where the class means are unconstrained (see Table 1). The relevant likelihood ratio statistic turned out to be nonsignificant. Hence, of all the models considered, the 3–class two-dimensional unfolding model with linear constraints on the stimulus coordinates is the most parsimonious model that fits the data well. The stimulus and class ideal point configuration obtained with this model is displayed in Figure 2, along with the true configuration from which the data were generated.

Insert Table 2 and Figure 2 about here

---

5. Illustrative Application

To illustrate the use of the latent class unfolding model, we apply it to a political science data set made available by the Department of Political Science of the University of Leiden and previously analyzed by Meulman and Verboon (1991). The data concern sympathy ratings for thirteen political parties made in the period 1979–1980 by 135 members of the Second Chamber of the Dutch Parliament. The parties that were included as stimuli are listed in Table 3 listed according to their rank order position on the left–right continuum. The 135 members of parliament (MP's) that participated in the study belonged to eight of the thirteen parties listed in Table 3. The exact parties to which the MP's belonged are given in the first column of Table 5.

Insert Table 3 about here

Following the model selection strategy suggested in Section 4, we first determined the appropriate number of latent classes by comparing the fit of a T–class model with no constraints on the class means with the fit of a $(T + 1)$–class unconstrained model using the Monte Carlo significance test procedure with $n = 500$. The results of these analyses are summarized in the upper part of Table 4. It can be inferred from the table that the appropriate number of classes is three. Subsequently, the 3–class two-dimensional unfolding model with a common additive constant α for all classes and the 3–class two-dimensional unfolding model with a separate additive constant α_t per class were fit to the data. The fit of these models was statistically compared with the fit of the 3–class unconstrained model by means of the Monte Carlo significance test (with $n = 500$). The results, reported in the lower part of Table 4, show that the 3–class two-dimensional unfolding model with a separate additive constant per class seems to fit the data equally well as the 3–class model with no constraints on the class means, while the 3–class two-dimensional unfolding model with a common additive constant must be rejected.

Insert Table 4 and Table 5 about here

The configuration obtained with this model is displayed in Figure 3. In the figure, the latent class ideal points are labeled A, B, and C. As argued in Section 3, it is possible to assign each subject to the class he or she is most likely to belong to on the basis of the posterior probabilities $h_{it}(\hat{\mathbf{A}}, \hat{\mathbf{B}}, \hat{\alpha}, \hat{\sigma}^2, \hat{\boldsymbol{\lambda}})$. Such a classification was carried out for the solution presented in Figure 3 and is summarized in Table 5. In Table 5, the rows refer to the political parties to which the respondents belonged, while the columns correspond to each of the three classes found with the latent class unfolding model. From the table, we can see that latent class A groups most of the MP's of the leftists parties PSP, PPR, PVA and D66. The members of the center parties ARP, KVP and CHU are mainly classified in latent class B. Latent class C, finally, groups the members of the VVD, which is a right-wing party. In Figure 3, an ellipse is drawn around the parties that correspond to each latent class on the basis of this classification.

Insert Figure 3 about here

The left–right dimension shows up on the horizontal axis except for the location of CPN. This somewhat unexpected location might be due to the fact that no communist MP's participated in the study. After the study was completed, the ARP, KVP, and CHU merged into a single party, called the CDA (Christian Democratic Appeal). Note that the members of these parties are classified on the basis of our analysis in the same latent class. It is also interesting to mention that the small left-wing parties (CPN, PPR, and PSP) form a single ecological party nowadays. The northwest–southeast direction distinguishes among three groups: the Christian democrats (ARP, KVP, and CHU), the liberals (D66 and VVD) and social democrats (PVA), and the extremist parties from both the left and the right. This distinction reflects one of the most enduring characteristics of Dutch politics: the Christian democrats have been in power for more than eighty years, but in changing coalitions with either the right (VVD) or the left

(PVA). The extremist parties have never been in power at all.

6. Discussion

In this paper, we developed a latent class unfolding model for single stimulus preference ratings. One of the advantages of the current approach is the possibility of testing the spatial unfolding model against the unconstrained latent class model for rating data. When carrying out an unfolding analysis by minimizing (3), it not possible to assess the validity of the spatial model. The assumptions of the latent class unfolding model are consistent with minimizing (3). In fact, they are less restrictive because independence is assumed only *within* the classes. In addition, the present approach leads to a much more parsimonious representation which should be easier to interpret in many cases. Also, if the model holds, one can expect more stable solutions.

One way to reduce the number of model parameters even more was elaborated in the paper when we discussed the possibility of imposing linear constraints on the stimulus coordinates through (8). It should be obvious that other ways of constraining the solution can easily be implemented within the present GEM framework. Most types of constraints currently used (such as linear constraints, centroid constraints, inequality constraints, etc.; see e.g., Heiser & Meulman, 1983), require a projection of $\tilde{\mathbf{B}}$ into some suitable subregion. This projection usually amounts to a weighted least squares problem (as for instance (43)). Likewise, constraints can be imposed on the latent class ideal points, although this might not be obvious in many applications since we do not know in advance the labeling of the classes. However, it is possible to include in a mixture distribution analysis subjects that are classified in advance, as discussed by McLachlan and Basford (1988). Such a hybrid analysis would enable us to establish a correspondence between the latent classes and a priori specified groups.

In the present approach, the equivalent of fitting an additional subject in a previously determined space (i.e., so-called external unfolding) simply amounts to classifying the new subject into a latent class on the basis of the posterior probabilities defined in (17). These posterior probabilities can also be used to define goodness-of-fit indices for individual subjects (see, Titterington, Smith, & Makov, 1985).

The application reported in Section 5 illustrates the usefulness of integrating a latent class point of view with a spatial unfolding model. A similar approach has been explored in the context of other multidimensional scaling models (De Soete & Winsberg, 1991; Winsberg & De Soete, in press). A different way of combining a classification approach with multidimensional scaling has been proposed by Heiser (1991b).

References

Carroll, J. D., Pruzansky, S., & Kruskal, J. B. (1980). CANDELINC: A general approach to multidimensional analysis of many-way arrays with linear constraints on parameters. *Psychometrika, 45*, 3–24.

Coombs, C. H. (1964). *A theory of data.* New York: Wiley.

de Leeuw, J. (1977). Applications of convex analysis to multidimensional scaling. In J. R. Barra, F. Brodeau, G. Romier, & B. Van Cutsem (Eds.), *Recent developments in statistics* (pp. 133–145). Amsterdam: North-Holland.

de Leeuw, J. (1988). Convergence of the majorization method for multidimensional scaling. *Journal of Classification, 5*, 163–180.

de Leeuw, J., & Heiser, W. (1977). Convergence of correction-matrix algorithms for multidimensional scaling. In J. C. Lingoes (Ed.), *Geometric representations of relational data* (pp. 735–752). Ann Arbor, MI: Mathesis Press.

de Leeuw, J., & Heiser, W. (1980). Multidimensional scaling with restrictions on the configuration. In P. R. Krishnaiah (Ed.), *Multivariate Analysis—V* (pp. 501–522). Amsterdam: North-Holland.

Dempster, A. P., Laird, N. M., & Rubin, D. B. (1977). Maximum likelihood estimation from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society, Series B, 39*, 1–38.

De Soete, G. (1990). A latent class approach to modeling pairwise preferential choice data. In M. Schader & W. Gaul (Eds.), *Knowledge, data and computer-assisted decisions* (pp. 103–113). Berlin: Springer-Verlag.

De Soete, G. (in press). Using latent class analysis in categorization research. In I. Van Mechelen, J. Hampton, R. Michalski, & P. Theuns (Eds.), *Categories and concepts: Theoretical views and inductive data analysis.* London: Academic Press.

De Soete, G., & DeSarbo, W. S. (1991). A latent class probit model for analyzing pick any/N data. *Journal of Classification, 8*, 45–63.

De Soete, G., & Winsberg, S. (1991). *A latent class vector model for analyzing preference ratings.* Paper presented at the 7th European Meeting of the Psychometric Society, Trier, Germany.

De Soete, G., & Winsberg, S. (in press). A Thurstonian pairwise choice model with univariate and multivariate spline transformations. *Psychometrika.*

Gibson, W. A. (1959). Three multivariate models: Factor analysis, latent structure analysis and latent profile analysis. *Psychometrika, 24*, 229–252.

Heiser, W. J. (1989). Order invariant unfolding analysis under smoothness restrictions. In G. De Soete, H. Feger, & K. C. Klauer (Eds.), *New developments in psychological choice modeling* (pp. 3–31). Amsterdam: North-Holland.

Heiser, W. J. (1981). *Unfolding analysis of proximity data.* Doctoral dissertation, University of Leiden, The Netherlands.

Heiser, W. J. (1987). The unfolding technique. In P. Legendre & L. Legendre (Eds.), *Developments in numerical ecology* (pp. 189–221). Berlin: Springer-Verlag.

Heiser, W. J. (1991a). A generalized majorization method for least squares multidimensional scaling of pseudodistances that may be negative. *Psychometrika, 56,* 7–27.

Heiser, W. J. (1991b). *Clustering of variables to optimize the fit of a low dimensional Euclidean model.* Paper presented at the 3rd meeting of the International Federation of Classification Societies, Edinburgh, Scotland.

Heiser, W. J., & Meulman, J. (1983). Constrained multdimensional scaling, including confirmation. *Applied Psychological Measurement, 22,* 139–167.

Hope, A. C. (1968). A simplified Monte Carlo significance test procedure. *Journal of the Royal Statistical Society, Series B, 30,* 582–598.

Kruskal, J. B. (1964). Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis. *Psychometrika, 29,* 1–27

Kruskal, J. B., Young, F. W., & Seery, J. B. (1973). *How to use KYST, a very flexible program to do multidimensional scaling and unfolding* (Technical Report). Murray Hill, NJ: Bell Laboratories.

Lazarsfeld, P. F., & Henry, R. W. (1968). *Latent structure analysis.* New York: Houghton Mifflin.

McLachlan, G. J. (1987). On bootstrapping the likelihood ratio test statistic for the number of components in a normal mixture. *Applied Statistics, 36,* 318–324.

McLachlan, G. J., & Basford, K. E. (1988). *Mixture models.* New York: Marcel Dekker.

Lord, F. M., & Novick, M. R. (1968). *Statistical Theories of Mental Test Scores.*

Reading, MA: Addison-Wesley.

Meulman, J. J., & Verboon, P. (1991). *Points of view analysis revisited: Fitting multidimensional structures to optimal distance components with cluster restrictions on the variables* (Technical Report RR–91–03). Leiden: University of Leiden, Department of Data Theory.

Takane, Y. (1976). A statistical procedure for the latent profile model. *Japanese Psychological Research, 18,* 82–90.

Titterington, D. M., Smith, A. F. M., & Makov, U. E. (1985). *Statistical analysis of finite mixture distributions.* Chichester: Wiley.

Wegman, E. J. (1972). Nonparametric probability density estimation. *Technometrics, 14,* 533–546.

Winsberg, S., & De Soete, G. (in press). A latent class approach to fitting the weighted Euclidean model, CLASCAL. *Psychometrika.*

Wolfe, J. H. (1970). Pattern clustering by multivariate mixture analysis. *Multivariate Behavioral Research, 5*, 329–350.

TABLE 1

Results of the Analyses of the Artificial Data

Analyses with No Constraints on the Class Means

No. of Classes (T)	Model df	Log Likelihood	Monte Carlo Significance Test of T versus $T+1$ Classes	
			Likelihood Ratio	Probability
1	17	-4449.3	942.4	< 0.01
2	34	-3978.1	1619.9	< 0.01
3	51	-3168.5	23.9	0.66

Analyses with 3–Class Unfolding Models

No. of Dimensions (R)	Linear Constraints	Model df	Log Likelihood	Monte Carlo Significance Test Against 3–Class Unconstr. Model	
				Likelihood Ratio	Probability
2	yes	39	-3172.4	7.8	0.81
2	no	21	-3186.0	35.0	0.27

TABLE 2

Design Matrix Used in the Artificial Data Example

Stimulus	Design Matrix					
1	1	0	0	1	0	0
2	0	1	0	1	0	0
3	0	0	1	1	0	0
4	-1	-1	-1	1	0	0
5	1	0	0	0	1	0
6	0	1	0	0	1	0
7	0	0	1	0	1	0
8	-1	-1	-1	0	1	0
9	1	0	0	0	0	1
10	0	1	0	0	0	1
11	0	0	1	0	0	1
12	-1	-1	-1	0	0	1
13	1	0	0	-1	-1	-1
14	0	1	0	-1	-1	-1
15	0	0	1	-1	-1	-1
16	-1	-1	-1	-1	-1	-1

TABLE 3

Parties included in the Political Science Study

Party	Description
CPN	Communists
PSP	Pacifistic socialists
PPR	Radical Christians
PVA	Social democrats
D70	Social democrats (economically conservative)
D66	Liberals (economically progressive)
ARP	Protestants (lower class)
KVP	Catholics
CHU	Protestants (upper- and middle class)
VVD	Liberals (economically conservative)
GPV	Very conservative Calvinists
SGP	Very conservative Calvinists
BP	Farmers Party

Note. The parties are listed from left to right

TABLE 4

Results of the Analyses of the Political Science Data Set

Analyses with No Constraints on the Class Means

No. of Classes (T)	Model df	Log Likelihood	Monte Carlo Significance Test of T versus $T+1$ Classes	
			Likelihood Ratio	Probability
1	14	-8025.5	758.1	< 0.01
2	28	-7646.5	210.9	< 0.01
3	42	-7541.1	109.0	0.17

Analyses with 3–Class Unfolding Models

No. of Dimensions (R)	Common α	Model df	Log Likelihood	Monte Carlo Significance Test Against 3–Class Unconstr. Model	
				Likelihood Ratio	Probability
2	yes	33	-7582.2	82.3	0.02
2	no	35	-7572.3	62.5	0.10

TABLE 5

Classification Based on the Posterior Probabilities

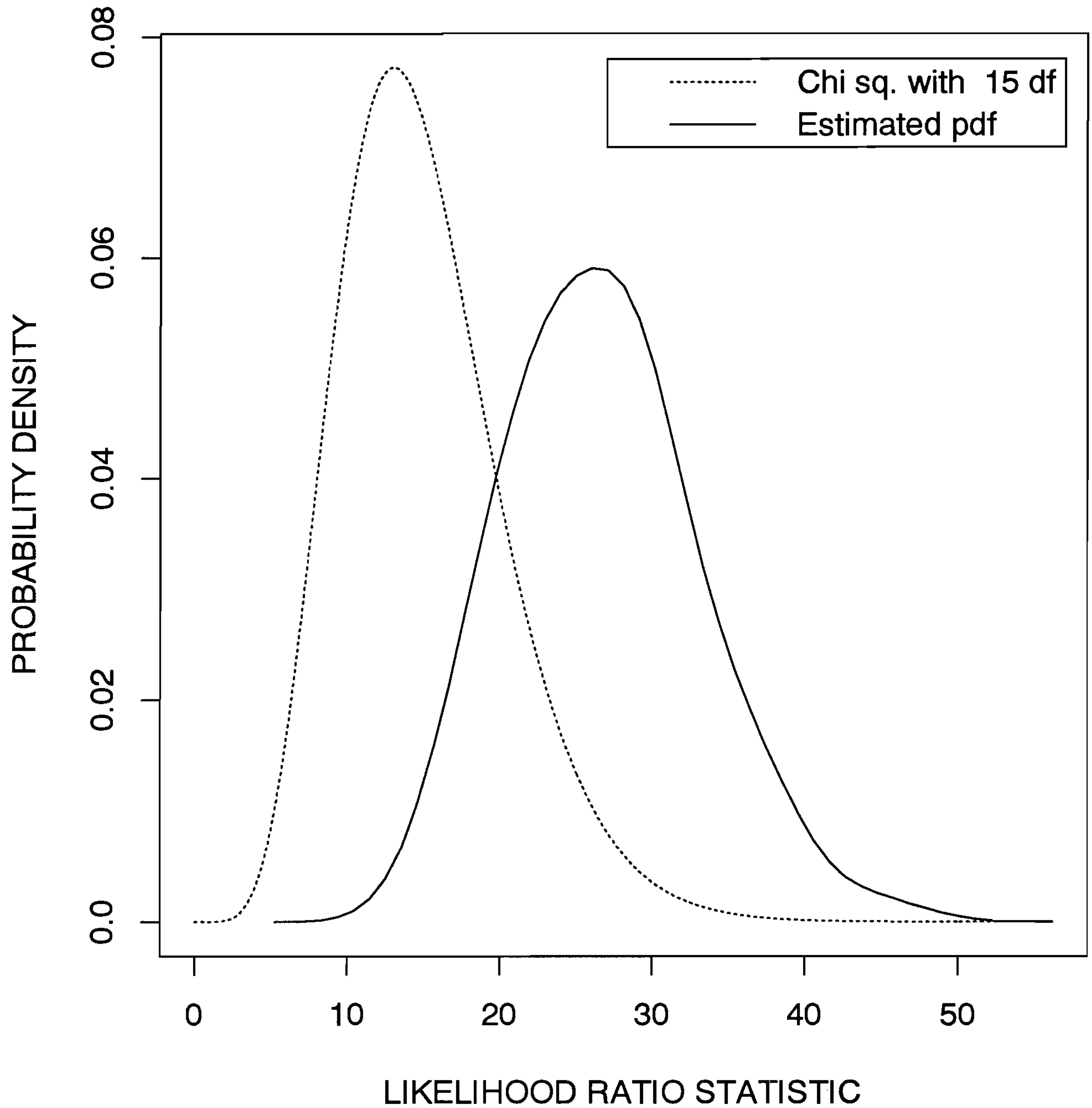
Party of Respondent	Latent Class A	B	C	No. of Respondents
PSP	1	–	–	1
PPR	3	–	–	3
PVA	52	1	–	53
D66	8	–	–	8
ARP	–	12	–	12
KVP	2	21	1	24
CHU	–	9	–	9
VVD	–	1	24	25
Total	66	44	25	135
λ_t	0.486	0.329	0.185	

Figure Captions

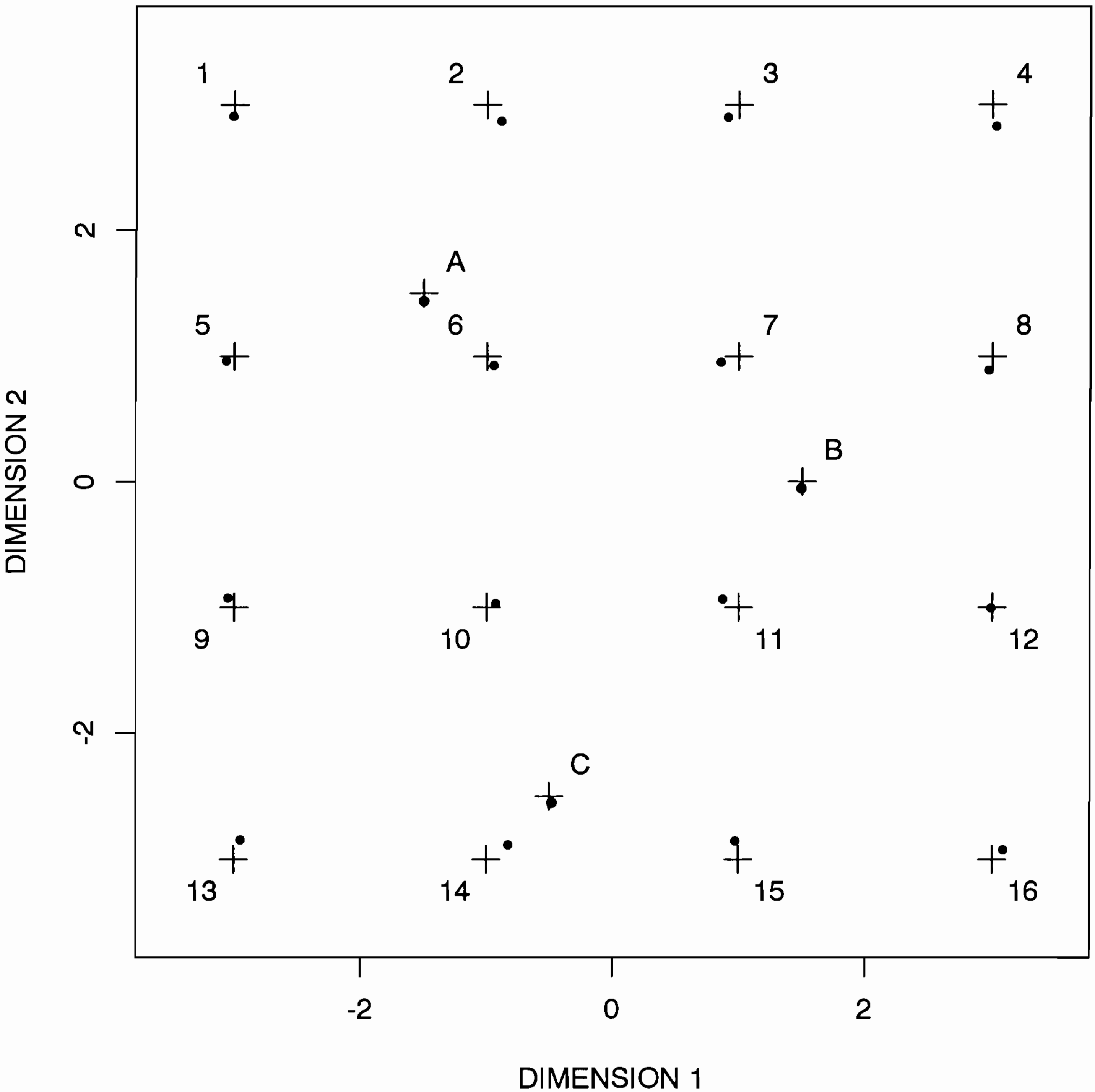
Figure 1. Estimated p.d.f. of the likelihood ratio statistic for testing an unconstraind two-class model versus an unconstrained three-class model. The dashed curve represent a chi-square distribution with degrees of freedom equal to the difference in degrees of freedom of the two models.

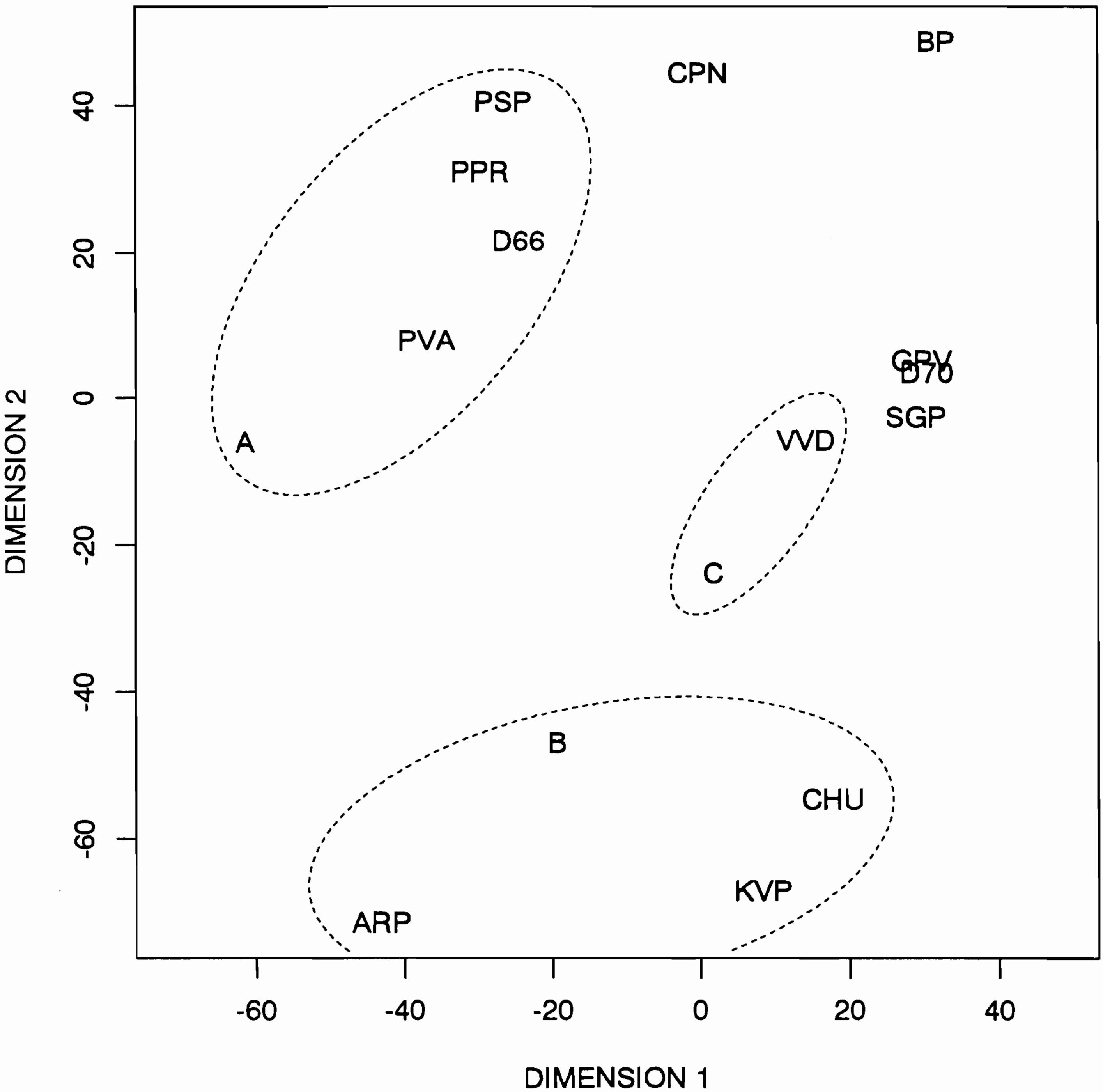
Figure 2. True and recovered solution for the artificial data example. The true points are indicated by crosses and the recovered points by dots.

Figure 3. Representation of the political science data according to a two-dimensional three-class unfolding model. The latent class ideal points are labeled A, B, and C.



Chi sq. with 15 df
Estimated pdf
PROBABILITY DENSITY
0.0
0.02
0.04
0.06
0.08
0
10
20
30
40
50
LIKELIHOOD RATIO STATISTIC


1
2
3
4
A
5
6
7
8
B
9
10
11
12
C
13
14
15
16
-2
0
2
DIMENSION 1
DIMENSION 2


BP
CPN
PSP
PPR
D66
PVA
A
GPV
D70
SGP
VVD
C
B
CHU
KVP
ARP
40
20
0
-20
-40
-60
-60
-40
-20
0
20
40
DIMENSION 2
DIMENSION 1