Classification and Related Methods of Data Analysis
H.H. Bock (Editor)
© Elsevier Science Publishers B.V. (North-Holland), 1988

PROBABILISTIC MULTIDIMENSIONAL SCALING OF PAIRED COMPARISONS DATA*)

Ingo BÖCKENHOLT and Wolfgang GAUL

Institute of Decision Theory and Operations Research,
Faculty of Economics,
University of Karlsruhe,
P.O. Box 69 80, D-7500 Karlsruhe, F.R.G.

In this paper two classes of probabilistic ideal point, respectively probabilistic vector, models are described and compared which - in different ways - account for the probabilistic nature of paired comparisons based choice behaviour. The main difference between both classes can be characterized in terms of the probabilistic utility specifications used. Whereas in the first class probabilistic ideal points or probabilistic vectors account for inconsistencies in the paired comparisons data, in the second class deterministic utility values yielded from non-probabilistic ideal points or vectors are superimposed by random error. A comparison is performed on the basis of theoretical aspects as well as empirical results.

1. INTRODUCTION

Inconsistencies in observed choice behaviour data have been a central motivation for researchers to develop probabilistic choice models. Especially, in paired comparisons data, intransitivities or circular triads may occur and a variety of approaches has been proposed, which try to model these inconsistencies. A review of these techniques based on criteria such as internal versus external analyses of pairwise choice data, unidimensional versus multidimensional representations, ideal point versus vector approaches etc., is given in e.g. Böckenholt/Gaul (1984, 1986), Bradley (1984) or De Sarbo/De Soete/Jedidi (1987).

In this paper we restrict our discussion to recently proposed, multidimensional scaling techniques, which - in different ways - account for the probabilistic nature of paired comparisons based choice behaviour. Essentially, two different classes of probabilistic ideal point models (PIPM), respectively probabilistic vector models (PVM), are described and compared the main difference of which can be explained in terms of different probabilistic utility specifications. In all cases, utility is described either by the distance between objects and ideal points (which is inversely related to utility) or by the projection of objects on preference vectors.

In the first class probabilistic ideal points, respectively probabilistic vectors, account for inconsistencies. The corresponding PVM is due to Carroll (1980) and De Soete/Carroll (1983) whereas the PIPM was independently developed by Böckenholt/Gaul (1986) and De Soete/Carroll/De Sarbo (1986).

Recently, De Sarbo/De Soete/Jedidi (1987) (and De Sarbo/Oliver/De Soete (1986)) have proposed a second class of multidimensional probabilistic choice models. The authors' approach is restricted to analyzing binary paired comparisons data and is based on a probabilistic choice theory similar to Thurstone's LCJ (Law

*) Research has been supported by the Deutsche Forschungsgemeinschaft.

of Comparative Judgement) Case V. More specifically, deterministic utility values yielded from non-probabilistic ideal points or preference vectors are superimposed by random error.

In this paper two aims are pursued. First of all - concerning more theoretical aspects - both classes of choice models are amalgamated into a unified probabilistic setting, in which the De Sarbo/De Soete/Jedidi/Oliver-approach is a special case. Hereafter - concerning more empirical aspects - both classes are compared in terms of empirical results yielded from the analysis of a data set known from previous research in choice theory.

2. PROBABILISTIC MULTIDIMENSIONAL SCALING

Since the landmark paper by Thurstone (1927) on unidimensional scaling of paired comparisons data, important progress has been made in probabilistic choice theory by developing models, which take into account the often multidimensional nature of choice objects. Researchers are especially interested in approaches, which allow for a subject-specific representation of choice behaviour, and it has become a common practice to realize such representations of judging subjects in terms of ideal points (see e.g. Bennett/Hays (1960), Carroll (1980), Coombs (1964), Schönemann (1970), Sixtl (1973), Zinnes/Griggs (1974)) or preference vectors (see e.g. Carroll (1980), Slater (1960), Tucker (1960)). The methodology presented below assumes the following relationship between ideal points, respectively preference vectors, and objects:

$s = 1,\dots,S$ subjects (or homogeneous groups of subjects; e.g. consumers), $l = 1,\dots,L$ objects (e.g. brands, print ads, package designs etc.) and $m = 1,\dots,M$ space dimensions are used for describing choice objects and judging subjects in an appropriate joint space. Ideal point coordinates $i_s^T = (i_{s1},\dots,i_{sM})$, respectively vector coordinates $v_s^T = (v_{s1},\dots,v_{sM})$, may be superimposed by random error, i.e. subject s in dimension m is described by

$$\begin{aligned} & i_{sm} + \varepsilon_{sm} \quad \text{or } v_{sm} + \varepsilon_{sm} \quad \text{with} \\ & \varepsilon_{sm} \sim N(0, \lambda_{sm}), \qquad \text{cov}\,(\varepsilon_{s_1 m_1}, \varepsilon_{s_2 m_2}) = 0, \end{aligned} \tag{1}$$

while object coordinates $x_l^T = (x_{l1},\dots,x_{lM})$ are deterministic (Here, T denotes transpose of a vector.).

We define a random weighted squared Euclidean distance between objects and ideal points by

$$U_{sl}^{i} = \sum_{m=1}^{M} w_{sm}\,(x_{lm} - (i_{sm}+\varepsilon_{sm}))^2 + \gamma_{sl}\,, \quad w_{sm} > 0, \tag{2a}$$

respectively a random scalar product between objects and preference vectors by

$$U_{sl}^{v} = \sum_{m=1}^{M} x_{lm}(v_{sm}+\varepsilon_{sm}) + \gamma_{sl}, \tag{2b}$$

where

$$\gamma_{s1} \sim N(0,\sigma_s), \qquad \mathrm{cov}(\gamma_{s_1 1_1}, \gamma_{s_2 1_2}) = 0, \tag{3}$$

describes additional random error not accounted for by the chosen dimensionality with

$$\mathrm{cov}(\gamma_{s_1 1}, \varepsilon_{s_2 m}) = 0. \tag{4}$$

Remember, that an ideal point i_s allows the identification of a most preferred perceptual space location of subject s for existing and/or new choice alternatives (e.g. products, package designs, print ad concepts etc.) under study, whereas a preference vector v_s indicates subjects's most preferred preference direction. The random error ε_{sm} is directly related to the coordinates of a deterministic ideal point i_s, respectively preference vector v_s, to account for within - subject inconsistencies (and across - subject inconsistencies when several subjects are considered to be replications of each other).

The error term γ_{s1} is introduced in a similar way as has been done by Thurstone (1927), McFadden (1976) and De Sarbo et al. (1986, 1987) to consider a variability within the data not accounted for by the dimensionality M of the joint space chosen.

Suppose a subject s is presented a pair of objects j,k, then object j will be preferred to k whenever

$$U^i_{sk} - U^i_{sj} > 0 \tag{5a}$$

or

$$U^v_{sj} - U^v_{sk} > 0 \tag{5b}$$

holds. With regard to the model assumptions (1), (2a,b), (3) and (4) the probability that the inequalities (5a), respectively (5b) are valid is given by

$$p_{sjk} = \begin{cases} \Pr(U^i_{sk} - U^i_{sj} > 0) = \Phi\left(\dfrac{\sum_{m=1}^{M} w_{sm}\,[2(x_{jm}-x_{km})i_{sm} + (x^2_{km}-x^2_{jm})]}{\left(4\sum_{m=1}^{M} w^2_{sm}\lambda^2_{sm}(x_{jm}-x_{km})^2 + 2\sigma^2_s\right)^{1/2}}\right) & (6a) \\[2ex] \Pr(U^v_{sj} - U^v_{sk} > 0) = \Phi\left(\dfrac{\sum_{m=1}^{M} (x_{jm}-x_{km})v_{sm}}{\left(4\sum_{m=1}^{M} \lambda^2_{sm}(x_{jm}-x_{km})^2 + 2\sigma^2_s\right)^{1/2}}\right) & (6b) \end{cases}$$

where Φ denotes the standard normal distribution function.

The parameters to be estimated in (6a,b) are the coordinates of the subjects' ideal points i_{sm}, respectively preference vectors v_{sm}, the coordinates of the objects' points x_{1m}, the weighting factors w_{sm} and the random error variances λ^2_{sm}, σ^2_s.

An extension of the above mentioned model equation via graded paired comparisons or via a reparametrization of the object, respectively subject, coordinates is straightforward and skipped here because of space limitations. Maximum likelihood parameter estimation procedures for special cases of equation (6a,b) along with several test statistics can be found in e.g. Arbuckle/Nugent (1973), Böckenholt/Gaul (1986), De Soete/Carroll/De Sarbo (1986), De Soete/Carroll (1983) and De Sarbo et al. (1986, 1987).

3. SPECIAL CASES

Various special cases derived from the methodology presented above can be obtained by assuming appropriate constraints with respect to the distribution assumptions (1) and/or (3). Here, we restrict ourselves to the two classes mentioned in the beginning of this paper. Therefore, it should, again, be stressed that for all approaches of the first class at least some kind of distribution assumption (1) is valid, while approaches of the second class can be described by distribution assumption (3) without consideration of (1), i.e. $\varepsilon_{sm} = 0$ (or $\lambda^2_{sm} \to 0$).

Concerning the first class, we receive the PIPM presented in Böckenholt/Gaul (1986), respectively the PVM described in De Soete/Carroll (1983), by fixing $\lambda^2_{sm} = 1, \forall$ s, m and $\sigma^2_s = \sigma^2, \forall$ s, while in the PIPM proposed by De Soete/Carroll/De Sarbo (1986) distribution assumption (1) but not (3) is used. Disregarding distribution assumption (3) simply means $\gamma_{sl} = 0$ (or $\sigma^2_s \to 0$) i.e. the term 2 σ^2_s in the denominator of (6a,b) has to be cancelled. An important characteristic of all approaches based on distribution assumption (1) is, that choice probability is not only a function of the difference in utility of the choice objects but also a function of the similarity or comparability between choice objects. This characteristic is apparent from (6a,b) and implies moderate stochastic transitivity, as has been proved in Halff (1976).

Using distribution assumption (3) and disregarding (1), i.e. omitting the random error term ε_{sm} (or $\lambda^2_{sm} \to 0$), we receive - together with the corresponding modified expressions (2a,b) and (6a,b) - the second class of approaches proposed by De Sarbo et al. (1986, 1987). Similar as Thurstone's (1927) LCJ Case V these approaches imply strong stochastic transitivity and don't account for similarity between choice alternatives. Mathematically, the terms $4 \sum_{m=1}^{M} w^2_{sm} \lambda^2_{sm} (x_{jm} - x_{km})^2$, respectively $4 \sum_{m=1}^{M} \lambda^2_{sm}(x_{jm}-x_{km})^2$, in the denominator of (6a,b) have to be cancelled. This means that the probability of choosing a choice object j from two alternatives (j,k) is modeled as a strictly increasing function of the difference in subjective utility of the choice objects j and k.

The question, however, is whether such a restriction is really necessary. Notice, that the x_{lm}-coordinates in the denominator terms have to be estimated anyhow. Under the assumption of equal weights (i.e. $w_{sm} = 1, \forall$ s,m), also used in De Sarbo/De Soete/Jedidi (1987), the λ_{sm}-parameters are the additional unknown quantities which have to be specified or estimated. Although the discussion of whether these parameters can be estimated within the used approach is not finished yet, one possibility is, of course, to specify these parameters, e.g. $\lambda_{sm} = 1 \, \forall$ s, m. Empirical implications yielded by this simple choice of the corresponding variance parameters will be illustrated in the next section.

4. EMPIRICAL RESULTS

A first empirical comparison of the two classes of PIPM, respectively PVM, was made on the basis of the Rummelhart/Greeno (1971) data, a well-known example from previous research in choice theory. In this study, 234 college students judged three groups of celebrities with respect to the question "With whom would you prefer to spend an hour of conversation?". The first group consisted of three politicians (Harold Wilson (HW), Charles De Gaulle (CD) and Lyndon B. Johnson (LJ)), the second group of three athletes (Johnny Unitas (JU), Carl Yastrzemski (CY) and A.J. Foyt (AF)) and the third group of three movie stars (Brigitte Bardot (BB), Elizabeth Taylor (ET) and Sophia Loren (SL)). Subjects were treated as replications within one group, so that $S = 1$ holds.

This data set is well suited to demonstrate to which extent a scaling technique is able to discover the perceived similarity of different choice objects, because it can be expected (and has been found out in a number of analyses; see e.g. Böckenholt/Gaul (1986), De Soete/Carroll/De Sarbo (1986) or Takane (1980)) that people belonging to the same professional group are more similar to each other than those in different groups.

Tab. 1 contains selected analyses of these data. The test values of Thurstone's unidimensional LCJ Case V model as well as the twodimensional PIPM, respectively PVM, approaches based on distribution assumption (3) with disregard of distribution assumption (1) indicate that these models have to be rejected. On the other side, the twodimensional PIPM, respectively PVM, approaches based on distribution assumption (1) with disregard of distribution assumption (3) lead to a significant fit improvement.

	Model specification				Test against null model			
	Dimensionality	Distrib. assumpt.	log L	Effective no. of parameters	χ^2	d.f.	p-value	AIC (-10000)
Null model	-	-	-5310.65	36	-	-	-	693.30
LCJ Case V	-	-	-5351.76	8	82.22	28	<0.001	719.52
Probabilistic ideal point model $w_{1m}=1, \forall\, 1, m$	2	(1) $\lambda_{1m}=1 \forall 1,m$	-5315.69	16	10.08	20	0.967	663.38
	2	(3)	-5351.74	17	82.18	19	<0.001	737.48
	2	(1)+(3) $\lambda_{1m}=1 \forall 1,m$	-5315.38	17	9.46	19	0.965	664.76
Probabilistic vector model	2	(1) $\lambda_{1m}=1 \forall 1,m$	-5317.68	16	14.06	20	0.827	667.36
	2	(3)	-5351.74	14	82.18	22	<0.001	731.48
	2	(1)+(3) $\lambda_{1m}=1 \forall 1,m$	-5317.61	17	13.92	19	0.834	669.22

Tab. 1: Summary of selected analyses on the Rummelhart/Greeno (1971) data

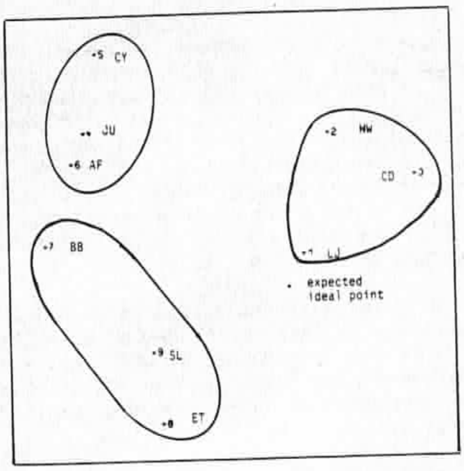

Fig. 1: Twodimensional solution of the Rummelhart/Greeno data according to the probabilistic ideal point model with distribution assumption (1) with $\lambda_{lm} = 1 \; \forall \; l,m$

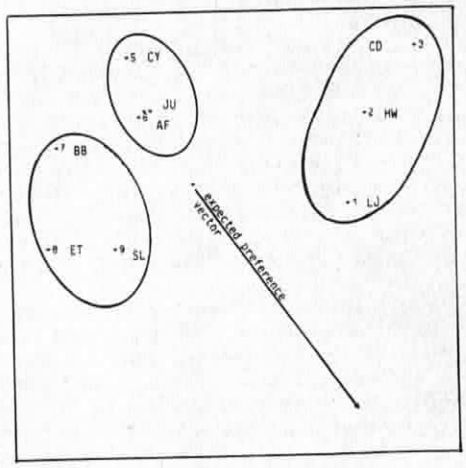

Fig. 2: Twodimensional solution of the Rummelhart/Greeno data according to the probabilistic vector model with distribution assumption (1) with $\lambda_{lm} = 1 \; \forall \; l,m$

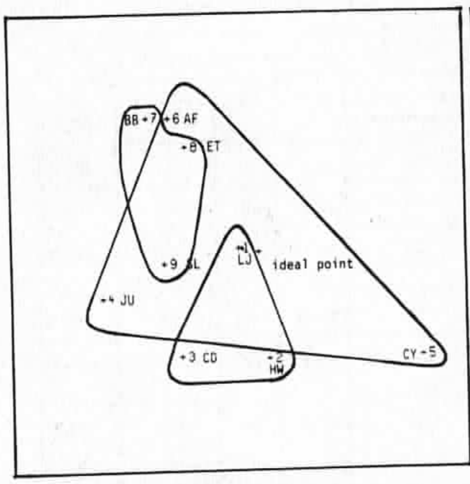

Fig. 3: Twodimensional solution of the Rummelhart/Greeno data according to the probabilistic ideal point model with distribution assumption (3)

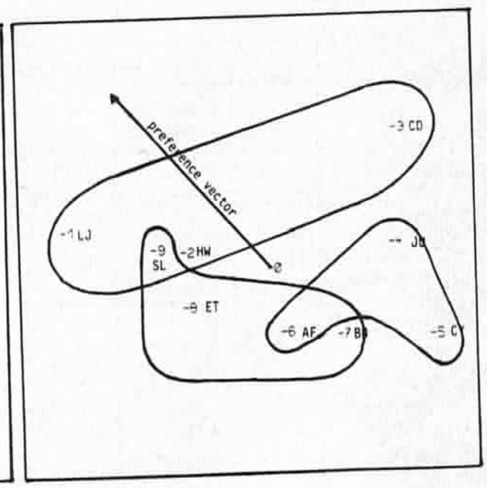

Fig. 4: Twodimensional solution of the Rummelhart/Greeno data according to the probabilistic vector model with distribution assumption (3)

This result is in so far remarkable as both distributions - for the different PIPM as well as for the PVM approaches - imply only slight differences with respect to the number of effective parameters which have to be estimated (see column 5 in Tab. 1). A simultaneous application of both distributions as well as the consideration of more than two dimensions (these results are not presented in Tab. 1 out of space limitations) lead to no significant fit improvements. Akaike's (1977) AIC - statistic indicates (the smaller the AIC, the better the model fits the data), that the two-dimensional PIPM based on distribution assumption (1) is most appropriate for representing the data.

Further hints which support the results of Tab. 1 are yielded by the two-dimensional graphical representations depicted in Fig. 1 to 4.

While there are only slight differences concerning the preference rank orders yielded from the different approaches (see Tab. 2), already a short glance at the figures shown reveals that the probabilistic choice models derived from distribution assumption (1) provide a more distinct cluster solution of the choice objects (see Fig. 1 and 2) than the probabilistic choice models based on distribution assumption (3) (see Fig. 3 and 4). In Fig. 1 and 2 each professional group forms a separate cluster (as expected), while not the same can be stated for Fig. 3 and 4. Therefore, one may conclude, that the first class of probabilistic choice models seems to be more appropriate in discovering the perceived similarity of different choice objects, at least for the S=1 group case.

Ideal point model	Fig. 1	1	2	9	3	8	6	4	7	5
Vector model	Fig. 2	1	2	9	3	8	6	4	7	5
Ideal point model	Fig. 3	1	9	2	8	3	6	4	7	5
Vector model	Fig. 4	1	9	2	8	3	6	4	7	5

Tab. 2: Rank orders yielded by the two-dimensional solutions depicted in Fig. 1 to 4

5. CONCLUSIONS

After having amalgamated two classes of choice models into a unified probabilistic setting and after having discussed theoretical aspects of various special cases, an empirical comparison of both classes was conducted. Although the results - concluded from the analysis of a data set well known from previous research in choice theory - indicate some important features for the special case S = 1, further research efforts on the basis of individual data have to be performed to allow a more comprehensive comparison of both classes. For example, it would be desirable to examine the performance of the different approaches on data sets with S(>1) groups of subjects (with replications within groups) or on data sets which are handled on an individual basis (with and without replications).

REFERENCES

[1] Akaike, H., On Entropy Maximization Principle, in: Krishnaiah, P. R. (ed.), Applications of Statistics. North Holland, Amsterdam, 1977, pp. 27 - 41.

[2] Arbuckle, J. and Nugent J. H., A General Procedure for Parameter Estimation for the Law of Comparative Judgement. British Journal of Mathematical and Statistical Psychology 26 (1973) 240 - 260.

[3] Bennett, J. F. and Hays, W. L., Multidimensional Unfolding: Determining the Dimensionality of Ranked Preference Data. Psychometrika 25 (1960) 27 - 43.

[4] Böckenholt, I. and Gaul W., A Multidimensional Analysis of Consumer Preference Judgements Related to Print Ads, in: Methodological Advances in Marketing Research in Theory and Practice. EMAC/ESOMAR Symposium, Copenhagen, 1984, pp. 83 - 105.

[5] Böckenholt, I. and Gaul, W., Analysis of Choice Behavior via Probabilistic Ideal Point and Vector Models. Applied Stochastic Models and Data Analysis 2 (1986) 202 - 226.

[6] Bradley, R. A., Paired Comparisons: Some Basic Procedures and Examples, in: Krishnaiah, P. R. and Sen, P. K., (eds.), Handbook of Statistics, Vol. 4. Elsevier Science Publishers, 1984, pp. 299 - 326.

[7] Carroll, J. D., Models and Methods for Multidimensional Analysis of Preferential Choice (or other Dominance) Data, in: Lantermann, E. D. and Feger H. (eds.), Similarity and Choice. Huber, Bern, 1986, pp. 234 - 289.

[8] Coombs, C. H., A Theory of Data. Wiley, New York, 1964.

[9] De Sarbo, W., Oliver, R. L. and De Soete, G., A Probabilistic Multidimensional Scaling Vector Model. Applied Psychological Measurement 10 (1986) 79 - 98.

[10] De Sarbo, W., De Soete, G. and Jedidi, K., Probabilistic Multidimensional Scaling Models for Analyzing Consumer Choice Behavior. Communication and Cognition 20 (1987) 93 - 116.

[11] De Soete, G. and Carroll J. D., A Maximum Likelihood Method for Fitting the Wandering Vector Model. Psychometrika 48 (1983) 553 - 566.

[12] De Soete, G., Carroll, J. D. and De Sarbo, W. S., The Wandering Ideal Point Model: A Probabilistic Multidimensional Unfolding Model for Paired Comparisons Data. Journal of Mathematical Psychology 30 (1986) 28 - 41.

[13] Halff, H. M., Choice Theories for Differentially Comparable Alternatives. Journal of Mathematical Psychology 14 (1976) 244 - 246.

[14] McFadden, B., Quantal Choice Analysis: A Survey. Annals of Economic and Social Measurement 5 (1976) 363 - 390.

[15] Rumelhart, D. L. and Greeno, J. G., Similarity Between Stimuli: An Experimental Test of the Luce and Restle Choice Models. Journal of Mathematical Psychology 8 (1971) 370 - 381.

[16] Schönemann, P. H., On Metric Multidimensional Unfolding. Psychometrika 35 (1970) 349 - 366.

[17] Sixtl, F., Probabilistic Unfolding. Psychometrika 38 (1973) 235 - 248.

[18] Slater, P., The Analysis of Personal Preferences. British Journal of Statistical Psychology 13 (1960) 119 - 135.

[19] Takane, Y., Maximum Likelihood Estimation in the Generalized Case of Thurstone's Model of Comparative Judgement. Japanese Psychological Research 22 (1980) 188 - 196.

[20] Thurstone, L. L., A Law of Comparative Judgement. Psychological Review 34 (1927) 273 - 286.

[21] Tucker, L. R., Intra-Individual and Inter-Individual Multidimensionality, in: Gulliksen H., and S. Messick (eds.), Psychological Scaling: Theory and Applications. Wiley, New York, 1960, pp. 155 - 167.

[22] Zinnes, J. L. and Griggs, R. A., Probabilistic Multidimensional Unfolding Analysis. Psychometrika 39 (1974) 327 - 350.