

# Fitting Unidimensional Choice Models with Nonmetric Multidimensional Scaling

Mark L. Davison  
University of Minnesota

Phillip K. Wood  
Search Institute

A class of unidimensional choice models is described. Thurstone's paired comparisons model, Case 5, and the Bradley-Terry-Luce model both fall into the class. A simple nonmetric method is presented for estimating scale values from choice data which satisfy any model in the class. In two examples, nonmetric

scale values are compared to Thurstone estimates. The scaling method is extended to permit estimation of scale values in a class of unidimensional ordered-category models, a class which includes the law of categorical judgment.

In choice studies, subjects are presented all possible pairs of a stimulus set. For each pair, the subjects are asked to choose that member which, in their judgment, contains more of some attribute specified by the researcher. The basic data from such an experiment are probabilities,  $p_{ij}$ , the proportion of trials in which stimulus  $j$  was chosen when the stimulus pair  $(i, j)$  was presented. These empirical probabilities are estimates of population values,  $\pi_{ij}$ , defined over a universe of trials.

A class of choice models that satisfies the conditions specified in Equations 1 to 3 will be considered. Let  $x_i$  and  $x_j$  be scale values for stimuli  $i$  and  $j$ , respectively, along the attribute under investigation. For the class of choice models to be considered,

$$\pi_{ij} = f(x_j - x_i) \quad [1]$$

Here  $f()$  is a monotone nondecreasing function such that

$$(x_j - x_i) > (x_{j'} - x_{i'}) \rightarrow f(x_j - x_i) \geq f(x_{j'} - x_{i'}) \quad [2]$$

and such that

$$f(0) = .50 \quad [3]$$

Equations 1 through 3 above describe a general class of models which Baird and Noma (1978) call generalized Fechnerian scaling models. Thurstone's (1927) law of comparative judgment, Case 5, is one example of such a model. In Thurstone's model,

---

APPLIED PSYCHOLOGICAL MEASUREMENT  
Vol. 7, No. 3, Summer 1983, pp. 333-340  
© Copyright 1983 Applied Psychological Measurement Inc.  
0146-6216/83/030333-08\$1.65

$$\pi_{ij} = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{(x_j - x_i)} e^{-t^2/2} dt \quad [4]$$

The Bradley-Terry-Luce (Bradley & Terry, 1952; Luce, 1959) model is another example from the general Fechnerian class:

$$\pi_{ij} = \frac{e^{(x_j - x_i)}}{1 + e^{(x_j - x_i)}} \quad [5]$$

Let  $\delta_{ij}$  be defined as follows:

$$\delta_{ij} = |\pi_{ij} - .50| \quad [6]$$

The next section will demonstrate that if the choice probabilities satisfy Equations 1 through 3, then  $\delta_{ij}$  is a monotone nondecreasing function of the distances  $|x_j - x_i|$  between stimuli along the judged attribute. That is, it will be shown, that if Equations 1 through 3 are satisfied, then

$$|x_j - x_i| > |x_{j'} - x_{i'}| \rightarrow \delta_{ij} \geq \delta_{i'j'} \text{ for all } i, i', j, j' \quad [7]$$

This means that the nonmonotone transformation of the choice probabilities  $\pi_{ij}$  is a dissimilarity measure meeting the fundamental assumption of nonmetric multidimensional scaling in one dimension. Given good estimates,  $p_{ij}$ , then it should be possible to estimate the scale values,  $x_i$  and  $x_j$ , using a nonmetric MDS analysis (Davison, 1983; Lingo, 1973; Kruskal & Wish, 1978; Kruskal, Young, & Seery, 1973; Schiffman, Reynolds, & Young, 1981; Takane, Young, & de Leeuw, 1977) in one dimension of the dissimilarity estimates,  $\delta_{ij} = |p_{ij} - .50|$ . The dissimilarity matrix,  $\Delta = \{\delta_{ij}\}$  should be unidimensional.

After proving the implication in Equation 7, two examples will be presented comparing the nonmetric, one-dimensional scale values to traditional Thurstone (1927) estimates.

#### Proof of Implication

To prove the implication in Equation 7, let  $a = x_j - x_i$  and  $b = x_{j'} - x_{i'}$ . Squaring the premise of Equation 7 yields

$$a^2 > b^2 \quad [8]$$

Subtracting  $b^2$  from both sides and expanding the resulting difference of squares yields

$$(a - b)(a + b) > 0 \quad [9]$$

Now consider four exhaustive conditions:

1.  $x_j \geq x_i$  and  $x_{j'} \geq x_{i'}$ ;
2.  $x_j < x_i$  and  $x_{j'} < x_{i'}$ ;
3.  $x_j \geq x_i$  and  $x_{j'} < x_{i'}$ ; and
4.  $x_j < x_i$  and  $x_{j'} \geq x_{i'}$ .

If the implication holds for these conditions, it holds in general.

*Condition 1:*  $x_j \geq x_i$  and  $x_{j'} \geq x_{i'}$ . Under this case,  $a + b > 0$  (except when  $x_j = x_i$  and  $x_{j'} = x_{i'}$ , in which case  $\delta_{ij} = \delta_{i'j'} = .50$ ). Dividing both sides in Equation 9 by  $a + b$  and then adding  $b$  to both sides yields

$$a > b \quad [10]$$

by Equation 2 and the definitions of  $f()$ ,  $a$ , and  $b$ .

$$\pi_{ij} \geq \pi_{i'j'} \quad [11]$$

Subtracting .5 from both sides,

$$\pi_{ij} - .5 \geq \pi_{i'j'} - .5 \quad [12]$$

By Equations 2 and 3,  $\delta_{ij} \geq .5$  and  $\delta_{i'j'} \geq .5$  are known. Therefore,

$$\delta_{ij} \geq \delta_{i'j'} \quad [13]$$

*Condition 2:*  $x_j < x_i$  and  $x_{j'} < x_{i'}$ . For this condition,  $a + b$  in Equation 10 is negative. Dividing both sides of Equation 10 by  $a + b$  and multiplying by  $-1$  yields

$$-a > -b \quad [14]$$

Proof proceeds as in Condition 1 except with  $i, i'$  reversed with  $j$  and  $j'$ , respectively, i.e.,

$$\delta_{ji} \geq \delta_{j'i'} \quad [15]$$

To prove  $\delta_{ji} = \delta_{ij}$  (and, by extension, that  $\delta_{j'i'} = \delta_{i'j'}$ ), recall that  $\pi_{ij} + \pi_{ji} = 1$ . It follows that  $\pi_{ij} = 1 - \pi_{ji}$ . Hence,

$$\begin{aligned} |\pi_{ij} - .5| &= |1 - \pi_{ji} - .5| \\ &= |-\pi_{ji} + .5| \\ &= |\pi_{ji} - .5| \end{aligned} \quad [16]$$

By the definition of  $\delta$  in Equations 6 and 16

$$\delta_{ij} = \delta_{ji} \quad [17]$$

By Equations 17 and 15

$$\delta_{ij} \geq \delta_{i'j'} \quad [18]$$

*Condition 3:*  $x_j \geq x_i$  and  $x_{j'} < x_{i'}$ . In this case, in Equation 9,  $a - b > 0$ . Proof proceeds as in Condition 1 except with roles of  $i'$  and  $j'$  reversed, i.e.,

$$\delta_{ij} \geq \delta_{j'i'} \quad [19]$$

which, by Equation 17 yields

$$\delta_{ij} \geq \delta_{i'j'} \quad [20]$$

*Condition 4:*  $x_j < x_i$  and  $x_{j'} \geq x_{i'}$ . In this case, in Equation 9,  $a - b < 0$ . Proof proceeds as in Condition 2 except with the roles of  $i'$  and  $j'$  again reversed, i.e.,

$$\delta_{ji} \geq \delta_{i'j'} \quad [21]$$

which by Equation 17, enables the desired conclusion,

$$\delta_{ij} \geq \delta_{i'j'} \quad [22]$$

In short, the proofs leading to Equations 13, 18, 20, and 22 show that in all four of the possible conditions, the implication in Equation 7 holds. If the choice probabilities satisfy Equations 1 through 3, then the dissimilarity measure,  $\delta_{ij}$ , will be a monotone increasing function of the distance between the two stimuli.

### Examples

#### Guilford Data

The first example is taken from Guilford (1954). Subjects were presented all possible pairs of nine vegetables. For each pair, the subject was to indicate which he/she preferred more. Table 1 shows the basic data matrix  $\mathbf{P}$  from the study. Each element of  $\mathbf{P} = \{p_{ij}\}$  is the proportion of subjects choosing vegetable  $j$  over vegetable  $i$ . A matrix of dissimilarity estimates was computed according to Equation 6. Matrix  $\Delta$  in Table 2 shows these dissimilarity estimates.

The bottom of Table 2 shows the nonmetric one-dimensional scale values as well as the Thurstone scale values for these nine stimuli. The nonmetric estimates were obtained from the ALSCAL program (Takane, Young, & de Leeuw, 1977; Young & Lewycky, 1979) specifying an ordinal level of analysis. All other parameters were set to the ALSCAL default values. For the one-dimensional solution, S-STRESS equaled .116, and STRESS equaled .153. The correlation between the Thurstone and nonmetric scale values was .99.

Despite this high correlation, two of the stimuli—string beans and peas—were not ordered the same way by the two sets of scale values. In looking at the original data in Table 1, it is not clear how these two stimuli should be ordered relative to each other. As compared to peas, string beans were preferred more frequently over turnips, beets, and asparagus. These comparisons suggest that string beans should receive a higher scale value than peas. On the other hand, as compared to string beans, peas were preferred more frequently over cabbage, carrots, spinach, and corn. When peas and string beans were presented together, peas were preferred 52.7% of the time. These latter data points suggest that peas should receive a higher scale value than string beans. When faced with these inconsistencies in the data, the Thurstone and nonmetric multidimensional scaling methods resolved those inconsistencies differently; the Thurstone procedure gave peas the higher scale value; the nonmetric algorithm gave string beans the higher scale value.

To determine whether this reversal might just reflect a nonmetric local minimum, the Thurstone scale values were used as a starting configuration for a new nonmetric one-dimensional solution. The

Table 1  
Matrix  $\mathbf{P}$  for the Vegetable Data

Vegetable	T	Cab	B	A	Car	Sp	SB	P	Co
Turnips (T)	—	.818	.770	.811	.878	.892	.899	.892	.926
Cabbage (Cab)	.182	—	.601	.723	.743	.736	.811	.845	.858
Beets (B)	.230	.399	—	.561	.736	.676	.845	.797	.818
Asparagus (A)	.189	.277	.439	—	.561	.588	.676	.601	.730
Carrots (C)	.122	.257	.264	.439	—	.493	.574	.709	.764
Spinach (Sp)	.108	.264	.324	.412	.507	—	.628	.682	.628
String Beans (SB)	.101	.189	.155	.324	.426	.372	—	.527	.642
Peas (P)	.108	.155	.203	.399	.291	.318	.473	—	.628
Corn (Co)	.074	.142	.182	.270	.236	.372	.358	.372	—

Reproduced by permission of J. P. Guilford

Table 2  
Matrix  $\Delta$  and the Scale Values for the Vegetable Data

Vegetable	T	Cab	B	A	Car	Sp	SB	P	Co
Turnips (T)									
Cabbage (Cab)	.318								
Beets (B)	.270	.101							
Asparagus (A)	.311	.223	.061						
Carrots (C)	.378	.243	.236	.061					
Spinach (Sp)	.392	.236	.176	.088	.007				
String Beans (SB)	.399	.311	.345	.176	.074	.128			
Peas (P)	.392	.345	.297	.101	.209	.182	.027		
Corn (Co)	.426	.358	.318	.230	.264	.128	.142	.128	
Scale Values									
Thurstone	-2.029	-.955	.684	.016	.265	.320	.846	.936	1.318
Nonmetric	-2.186	-.727	-.642	-.172	.286	.555	.991	.744	1.151

new nonmetric scale values ordered the stimuli in the same way as the nonmetric scale values in Table 2.

#### Ayres Data

The second example is the classic Ayres handwriting data (Gulliksen & Tukey, 1958). Subjects judged all possible pairs of nine handwriting samples. For each pair, the subjects selected that sample which was, in their opinion, of higher quality. Table 3 shows the choice probabilities in matrix  $P$  and the dissimilarities estimated according to Equation 6. The ALSCAL (Young & Lewycky, 1979) analysis in one dimension employed the same options as the first example and yielded the scale values shown at the bottom of Table 4. S-STRESS equaled .240 and STRESS equaled .195. The correlation between the Thurstone scale values and the nonmetric solution was .98.

Table 3  
Matrix  $P$  for the Handwriting Data

Handwriting Sample	50a	50b	50c	70a	70b	70c	80a	80b	80c
50a	-	.52	.67	.95	.99	.98	.99	.97	.94
50b	.48	-	.60	.85	.95	.96	.98	.98	.95
50c	.33	.40	-	.76	.78	.92	.91	.86	.96
70a	.05	.15	.24	-	.76	.87	.95	.79	.78
70b	.01	.05	.22	.24	-	.74	.80	.52	.71
70c	.01	.04	.08	.13	.26	-	.59	.26	.56
80a	.01	.02	.09	.05	.20	.41	-	.15	.31
80b	.03	.02	.14	.21	.48	.74	.85	-	.61
80c	.06	.05	.04	.22	.29	.44	.69	.39	-

The authors thank Harold Gulliksen and *Psychometrika* for permission to reproduce this table.

Again, despite the high correlation, the stimuli are not ordered in the same way by the two sets of scale values. The violations concern stimuli 70C and 80C. If 70C and 80C were removed, the remaining stimuli would be ordered the same way by the two procedures. Inspection of Table 3 shows that the position of 70C and 80C relative to the position of 80B is ambiguous in much the same way that the ordering of string beans and peas was ambiguous in the previous example.

To determine whether the location of 70C and 80C in the nonmetric solution might simply represent a local minimum, the Thurstone scale values were used as a starting configuration for a new nonmetric solution. The ordering of the stimuli was that of the nonmetric scale values in Table 4. The new nonmetric scale values ordered the stimuli in exactly the same way as the nonmetric scale values in Table 4.

Since the above analysis assumes that the data,  $\delta_{ij}$ , satisfy a one-dimensional distance model, researchers may wish to use axiomatic tests to assess whether the model assumptions are reasonable. Many axiomatic tests in the literature (Davison, 1983; Tversky & Gati, 1982) do not apply to the present problem either because they assume metric data (data related to distances by a linear function) or because they presume a multidimensional stimulus set. The transitivity axiom does apply. According to the transitivity axiom (Tversky & Gati, 1982), if for four stimuli ( $a, b, c, d$ )  $\delta_{ac} > \delta_{ab}$ ,  $\delta_{bc}$  and  $\delta_{bd} > \delta_{bc}$ ,  $\delta_{bd}$ ; then  $\delta_{ad} > \delta_{ab}$ ,  $\delta_{bd}$  and  $\delta_{ad} > \delta_{ac}$ ,  $\delta_{cd}$ .

Both of the data sets contain violations of this axiom. For instance, in the first example,  $\delta_{Sp,P} > \delta_{Sp,Sb}$ ,  $\delta_{Sb,P}$  and  $\delta_{Sb,Co} > \delta_{Sb,P}$ ,  $\delta_{P,Co}$ ; but  $\delta_{Sp,Co} \not> \delta_{Sp,Sb}$ ,  $\delta_{Sb,Co}$  and  $\delta_{Sp,Co} \not> \delta_{Sp,P}$ ,  $\delta_{P,Co}$ . In the second example, it seems impossible to find an ordering of the four stimuli (50b, 70c, 80b, and 80c) for which the data will satisfy the axiom. These violations help explain why the data do not fit the model perfectly. In addition, it appears that the Thurstone and nonmetric procedures resolved these inconsistencies differently; hence, their scale value estimates order the stimuli differently.

#### Extensions to Category Models

In the category judgment task, a subject must sort stimuli into unidimensional ordered categories, each of which corresponds to an interval along a unidimensional attribute specified by the experimenter. Subjects are instructed to place stimuli with more of the attribute in higher categories.

If there are  $n+1$  categories, there are then  $n$  boundaries between categories. Probabilistic models can be constructed for category judgment data which parallel the choice models described by Equations

Table 4  
Matrix  $\Delta$  for the Handwriting Data

Handwriting Sample	50a	50b	50c	70a	70b	70c	80a	80b	80c
50a									
50b	.02								
50c	.17	.10							
70a	.45	.35	.26						
70b	.49	.45	.28	.26					
70c	.48	.46	.42	.37	.24				
80a	.49	.48	.41	.45	.30	.09			
80b	.47	.48	.36	.29	.02	.24	.35		
80c	.44	.45	.46	.28	.21	.06	.19	.11	
Scale Values									
Thurstone	-1.616	-1.367	-.838	-.219	.443	.954	1.313	.515	.818
Nonmetric	-1.631	-1.340	-.865	-.331	.654	.878	1.133	.945	.557

1 through 3. Here  $\pi_{ij}$  refers to the probability that stimulus  $i$  will be placed in a category below boundary  $j$ . From subjects' categorizations of stimuli, empirical estimates of these probabilities  $p(i, j)$  can be obtained.

Let  $x_i$  be the location of stimulus  $i$  along the attribute being scaled, and let  $x_j$  refer to the location of category boundary  $j$ . For the class of categorization models considered here, the categorization probabilities,  $\pi_{ij}$ , must satisfy Equations 1 through 3. The law of categorical judgment (Torgerson, 1958) describes categorization probabilities which satisfy those equations:

$$\pi_{ij} = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{(x_j - x_i)} e^{-t^2/2} dt \quad [23]$$

Let  $\delta_{ij}$  be defined as in Equation 6. Using the same basic steps given in Equations 8 through 22, it can be shown that if the categorization probabilities satisfy Equations 1 through 3, then

$$|x_j - x_i| > |x_{j'} - x_{i'}| \rightarrow \delta_{ij} \geq \delta_{i'j'} \text{ for all } i, i', j, j' \quad [24]$$

In short, the dissimilarity measure,  $\delta_{ij}$ , will be a monotone function of distances between stimuli and category boundaries. Any nonmetric scaling algorithm (Kruskal et al., 1973; Lingoes, 1973; Takane et al., 1977) could be justifiably used to estimate stimulus scale values and category boundary locations, given good estimates of the categorization probabilities.

### Discussion

The approach described above is not the only way to estimate scale values from choice probabilities which satisfy the very general assumptions of Equations 1 through 3. Winsberg and Ramsey (1981) have suggested a method which provides scale value estimates from the choice probabilities directly and avoids the necessity of first transforming the data as in Equation 6. The Winsberg and Ramsey (1981) method uses a spline technique, which is as yet unfamiliar to most researchers and for which the required computer programs are often unavailable. Nonmetric multidimensional scaling programs are more readily available and the method is more familiar to many investigators.

Other authors have discussed nonmetric scaling of binary choice data derived as in Equation 6. Guttman (cited but not referenced in Carroll, 1980) proposed  $\delta_{ij} = |\pi_{ij} - .50|$  as a dissimilarity measure on which to base a multidimensional scaling of stimuli. Carroll (1980) dismissed the measure as too ad hoc, although adding that it did lead to meaningful configurations in some cases. Like Carroll, the authors of this paper offer no justification for using  $\delta_{ij}$  as a basis for a multidimensional scaling, but the proof above shows under what conditions it should provide a reasonable basis for a unidimensional scaling of stimuli.

Heiser (1981) briefly suggested  $\delta_{ij}$  as a measure on which to base a unidimensional scaling, although he provided no justification for the dissimilarity measure. Heiser worried, however, about local minima problems which might be encountered. Indeed, in research that simulated data, local minima problems have occurred more frequently with solutions in one dimension than for any other dimensionality. In real data applications, such as that of Young (1970) and the ones above, the unidimensional nonmetric scaling approach has often worked quite well. Future improvements in nonmetric multidimensional scaling algorithms may reduce or even eliminate the local minima problem in the unidimensional case. With existing algorithms, the local minima problem can be minimized by using a rational starting configuration or multiple starts—at least if choice data is being fit.

Category rating data pose a different problem. If the data come from a pairwise preference task, then the proximity matrix is square and symmetric. For each pair of objects (stimuli), there is an estimate of the distance between them. On the other hand, if the data come from a category rating task, then the proximity matrix  $\Delta$  will be a rectangular array with one row for each stimulus and one column for each category boundary. It contains ordinal estimates of the distance between each stimulus and each boundary but not estimates of the distances between stimuli or between boundaries. Hence, the statistical estimation problem resembles the unfolding problem in that there are two sets of objects (stimuli and boundaries) and one has estimates of the interset distances but not the within-set distances. Computational problems, such as local minima, have been more severe in the analysis of rectangular arrays than in the analysis of the more traditional square symmetric proximity data. It will take further research to determine the conditions under which nonmetric multidimensional scaling can successfully recover the stimulus and category boundary locations.

Use of nonmetric algorithms to fit choice data has three advantages, two of which were mentioned above. First, this approach requires only minimal assumptions about the monotone function  $f$  in the choice model. Second, it relies on nonmetric programs which are readily available to many researchers, becoming more so as time passes. Finally, extreme choice probabilities, those near 0 and 1, cannot be readily handled in many algorithms for computing scale values. Such extreme choice probabilities pose less of a problem in a nonmetric multidimensional scaling approach.

### References

- Baird, J. C., & Noma, E. *Fundamentals of scaling and psychophysics*. New York: Wiley, 1978.
- Bradley, R. A., & Terry, M. E. Rank analysis of incomplete block designs: I. The method of paired comparisons. *Biometrika*, 1952, 39, 324–345.
- Carroll, J. D. Models and methods for multidimensional analysis of preferential choice or other dominance data. In E. D. Lanterman & H. Feger (Eds.), *Similarity and choice*. Bern, Switzerland: Hans Huber, 1980.
- Davison, M. L. *Multidimensional scaling*. New York: Wiley, 1983.
- Guilford, J. P. *Psychometric methods*. New York: McGraw-Hill, 1954.
- Gulliksen, H., & Tukey, J. Reliability for the law of comparative judgment. *Psychometrika*, 1958, 23, 95–110.
- Heiser, W. J. *Unfolding analysis of proximity data*. Leiden, The Netherlands: Reprodienst Psychologie, 1981.
- Kruskal, J. B., & Wish, M. *Multidimensional scaling*. Beverly Hills CA: Sage Publications, 1978.
- Kruskal, J. B., Young, F. W., & Seery, J. B. *How to use KYST, a very flexible program to do multidimensional scaling and unfolding*. Unpublished manuscript, Bell Laboratories, 1973.
- Lingoes, J. C. *The Guttman-Lingoes nonmetric program series*. Ann Arbor MI: Mathesis Press, 1973.
- Luce, R. D. *Individual choice behavior*. New York: Wiley, 1959.
- Schiffman, S. S., Reynolds, M. L., & Young, F. W. *Introduction to multidimensional scaling*. New York: Academic Press, 1981.
- Takane, Y., Young, F. W., & de Leeuw, J. Nonmetric individual differences multidimensional scaling: An alternating least squares method with optimal scaling features. *Psychometrika*, 1977, 42, 7–67.
- Thurstone, L. L. A law of comparative judgment. *Psychological Review*, 1927, 34, 273–286.
- Torgerson, W. S. *Theory and methods of scaling*. New York: Wiley, 1958.
- Tversky, A., & Gati, I. Similarity, separability, and the triangle inequality. *Psychological Review*, 1982, 89, 123–154.
- Winsberg, S., & Ramsey, J. O. Analysis of pairwise preference data using integrated B-splines. *Psychometrika*, 1981, 46, 171–186.
- Young, F. W. Nonmetric scaling of line lengths using latencies, similarity, and same-different judgments. *Perception and Psychophysics*, 1970, 8, 363–369.
- Young, F. W., & Lewycky, R. *ALSCAL 4 User's Guide* (2nd ed.). Chapel Hill NC: Data Analysis and Theory Associates, 1979.

### Author's Address

Send requests for reprints or further information to Mark L. Davison, Department of Educational Psychology, University of Minnesota, 178 Pillsbury Drive S.E., Minneapolis MN 55455, U.S.A.