PSYCHOMETRIKA–VOL. 46, NO. 2.
JUNE, 1981

ANALYSIS OF PAIRWISE PREFERENCE DATA USING INTEGRATED B-SPLINES

SUZANNE WINSBERG

UNIVERSITÉ DE MONTRÉAL

JAMES O. RAMSAY

MCGILL UNIVERSITY

Pairwise preference data are represented as a monotone integral transformation of difference on the underlying stimulus-object or utility scale. The class of monotone transformations considered is that in which the kernel of the integral is a linear combination of B-splines. Two types of data are analyzed: binary and continuous. The parameters of the transformation and the underlying scale values or utilities are estimated by maximum likelihood with inequality constraints on the transformation parameters. Various hypothesis tests and interval estimates are developed. Examples of artificial and real data are presented.

Key words: monotone transformation.

The purpose of this communication is to propose a general method of scaling pairwise preference data that may be used without prior knowledge of the monotone function which best represents the relationship between a datum and the process giving rise to it. We represent choice data as an empirically determined monotone transformation of difference on the underlying stimulus-object or utility scale. The approach here is similar to the B-spline procedure used by Winsberg and Ramsay [1980] for monotone transformation of all or some of the variables in the regression model so as to maximize additivity. Other general methods have been proposed for analyzing choice data with general monotone functions [Kruskal, 1965; Young, de Leeuw & Takane, 1976; Takane, Young & de Leeuw, 1980], but a feature of this approach is that the response function, although empirically determined, can have the properties of a cumulative distribution function, and in any case is differentiable. This method should also be useful in a wide variety of problems where the form of a response function is unknown, and in particular applications of this approach to additive models is straightforward.

We are considering two types of pairwise preference data. We define binary data as observations of frequencies X_{ij} indicating the number of times stimulus j is chosen over stimulus i. We shall denote the number of stimuli by n and the number of replications as N. Various models have been proposed to analyze binary data. These models, all of which specify a particular response function, include those of Thurstone [1927], Bradley & Terry [1952], Ford [1957], and Luce [1959]. We define continuous data as observations of pairwise preference strengths X_{ij} indicating the degree to which stimulus j is preferred to stimulus i by a given subject or group of subjects. Bechtel [1976] has used a linear transformation of the pairwise preference strengths to develop a unidimensional scale of utility.

Our problem is to find the set of stimulus or utility values s_j $j = 1, \ldots, n$ and a transformation F of differences $s_j - s_i$ on the underlying utility difference scale, such that an

This investigation was supported in part by a research grant from the Natural Sciences and Engineering Research Council Canada.

Requests for reprints should be sent to S. Winsberg, Section Mesure et Évaluation, Faculté des Sciences de l'Éducation, Université de Montréal, Montréal, P. Québec, Canada. H3C 3T3.

0033-3123/81/0600-3013$00.75/0

objective function Q is optimized with respect to the s_j and F subject to monotonicity constraints on the transformation. Other constraints may also be required. For example, one may be looking for a transformation whose derivative is symmetric about a given point such as the origin on the difference scale.

Integrated B-Spline Monotone Transformations

We propose that F be a monotone transformation of the difference $(s_j - s_i)$ for both binary and continuous data. In the former case $F(s_j - s_i)$ is the probability that stimulus-object j is chosen over stimulus-object i and in the latter case $F(s_j - s_i)$ is the degree to which object j is preferred over object i. Consider monotone transformations of the form

$$F_{ij} = F(s_j - s_i) = F_0 + \int_{x_0}^{s_j - s_i} v(x)\, dx \tag{1}$$

where s_i and s_j are scale values assigned to objects i and j, x_0 is a fixed point on the difference scale such that

$$x_0 \leq \min_{ij}(s_j - s_i),$$

and F_0 is a constant.

For binary data F_0 must be nonnegative and $0 \leq F_{ij} \leq 1$, while for continuous data the constant F_0, which may be negative, is needed to preserve the natural origin of the scale. An advantageous feature of an integral transformation is that the monotone function $F(s_j - s_i)$ will be smooth even when $v(x)$ is not. Moreover the monotonicity constraints will be satisfied provided that $v(x)$ is nonnegative. We are therefore looking for a family of nonnegative functions to define $v(x)$, which are extremely flexible in shape and depend upon a limited number of parameters.

Spline functions have many useful properties and have found much use as approximating functions since they are very flexible and do not depend on an excessive number of parameters. Moreover B-splines have the added feature of being nonnegative, and it is these that we have chosen to define $v(x)$. Spline functions are piecewise polynomials of a given degree joined at values of their arguments called knots, and it is useful to refer to their order k defined as the maximum degree of the polynomials plus one. Splines or their derivatives may have discontinuities at the knots. The choice of the number and location of the knots can be made a priori or post hoc.

Let T knots forming a nondecreasing sequence span a specific interval $\{x_l, x_u\}$. Then any spline of order k on this interval can be represented as an $m = T + k$ term linear combination of B-splines, or *basis splines* of that order. This family forms a basis for spline functions which is well-conditioned evaluated easily, and nonnegative. Curry and Schoenberg [1947, 1966] introduced B-splines and demonstrated their positivity. Computational algorithms for easy evaluation were developed by Cox [1972] and de Boor [1972], whose book [de Boor, 1978] contains a good description of the properties of B-splines.

The choice of knot sequence determines the desired amount of smoothness at any point: the number of continuity conditions at the point plus the number of knots at the point is always equal to k. For example if there are no knots at a point, the function and its first $(k - 1)$ derivatives are continuous. If there is one knot at a point, the function and its first $(k - 2)$ derivatives are continuous. If there are k knots at a point the function is discontinuous at that point. In addition to the $m - k$ interior knots $t_q (q = k + 1, \ldots, m)$ there are k initial and k final knots which must not lie within the domain of interest. In many cases a convenient choice is to let the first k knots equal $\min\{x\}$ and the last k knots equal $\max\{x\}$. In this way one imposes no continuity conditions at the end points of the interval of interest.

The curve tends to be insensitive to knot choice provided there are no discontinuities and the curve is smooth in the sense of a small bound on the modulus of its second derivative. In any case we will only attempt to apply B-splines with fixed knots in this paper. Nevertheless techniques exist for improving knot choice [de Boor, 1978, Ch. XII; Ramsay, 1977; de Boor & Rice 1968, Note 1]. We will consider sensitivity to knot choice in a small way in this paper and more seriously in future work.

In order to be parsimonious with respect to the number of transformation parameters and to obtain a reasonably good fit, two guidelines should be used for the a priori placement of knots: first, have as few knots as possible, while ensuring that there are at least four or five values of x between any two knots; and second, locate more knots where the nonlinearity of the function is expected to be more pronounced. Afterwards a fairly crude technique may be used to improve knot choice.

Once the knots have been fixed the value of a B-spline function of order k can be generated from the following recursion formula [de Boor, 1972]:

$$\begin{aligned}
B_{q1}(x) &= 1 \qquad \text{for } t_q \leq x \leq t_{q+1} \\
B_{q1}(x) &= 0 \qquad \text{otherwise} \\
B_{qk}(x) &= \frac{(x - t_q)}{(t_{q+k-1} - t_q)} B_{q,k-1}(x) + \frac{(t_{q+k} - x)}{(t_{q+k} - t_{q+1}} B_{q-1,k-1}(x) \qquad (2) \\
&\qquad \text{for } (t_{q+k} - t_{q+1}), (t_{q+k-1} - t_q) > 0 \\
B_{qk}(x) &= 0 \qquad \text{otherwise.}
\end{aligned}$$

It should be noted that $B_{qk}(x)$ is positive for $t_q \leq x \leq t_{q+k}$ and zero otherwise. It follows that only k of the B-splines are positive at any point $x, t_k \leq x \leq t_{m+1}$. For example, the five order two B-splines which form the basis for the knot sequence $\{0, 0, .2, .5, .8, 1.0, 1.0\}$ are illustrated in Figure 1 (above). Note that only two B-splines are positive between any pair of knots. Since only k of the B-splines are positive in any interval we may write

$$v(x) = \sum_{q=p-k+1}^{p} a_q B_{qk} \quad \text{for} \quad t_p \leq x \leq t_{p+1}. \qquad (3)$$

An example of a particular linear combination of order two B-splines defined by the knot sequence 0, 0, .2, .5, .8, 1.0, 1.0 is shown in Figure 1. Although the function $v(x)$, shown in Figure 1, is not smooth the integral $\int v(u)\, du$ also shown in Figure 1 is quite smooth.

Of particular interest in many applications is the ease with which linear combinations of B-splines can be integrated and differentiated [de Boor, 1978, Ch. X]. The integral of a linear combination of B-splines of order k is a linear combination of B-splines of order $k + 1$; the derivative of a linear combination of B-splines of order k is a linear combination of B-splines of order $k - 1$. Thus, for $t_1 \leq x \leq t_L$,

$$\int_{t_1}^{x} \left[\sum_{q=1}^{m} a_q B_{qk}(u) \right] du = \sum_{q=1}^{L=1} C_q B_{q,k+1}(x), \qquad (4)$$

where $C_q = \sum_{p=1}^{q} a_p(t_{p+k} - t_p)/k$, and

$$\frac{d}{dx}\left[\sum_{q=1}^{L} a_q B_{qk}(x) \right] = (k-1) \sum_{q=1}^{L+1} w_q B_{q,k-1}(x),$$

where $w_q = (a_q - a_{q-1})/(t_{q+k-1} - t_q)$ and $a_0 = a_{L+1} = 0$.

Finally B-splines are invariant with respect to scale transformation provided that the knot sequence is similarly transformed. In conclusion splines are more flexible than parametric families such as the normal and the logistic, more parsimonious and almost as

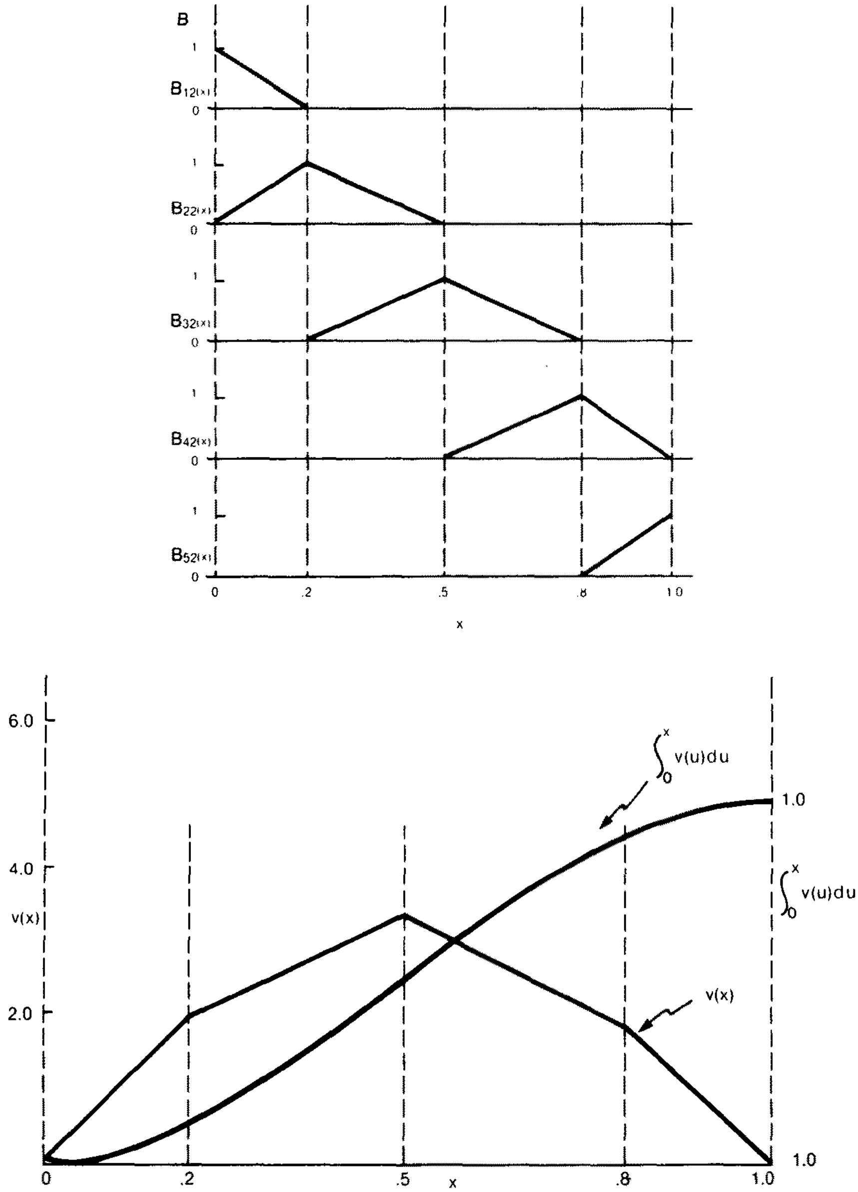

FIGURE 1
Upper: The order two B-splines for the knot sequence $\{0, 0, .2, .5, .8, 1.0, 1.0\}$.
Lower: A linear combination $v(x)$ of the B-splines shown at the left, and the integral $\int_0^x v(u)\,du$, where $a_1 = 0.03$, $a_2 = 1.96$, $a_3 = 3.40$, $a_4 = 1.96$ and $a_5 = 0.02$. Vertical dashed lines indicate knot positions.

flexible as monotone functions without further restrictions, more flexible than monotone polynomials of low degree, and more parsimonious than and just as flexible as polynomials of high degree.

Point Estimation, Interval Estimation, and Hypothesis Tests

Two types of choice data are under consideration: binary data and continuous data. A similar approach to the analysis of both types of data will be presented first, and then aspects of the approach particular to each type of data will be discussed. In fitting integrated B-splines, we assume that the choice data are independently distributed with probability density function

$$\prod_{i=1}^{n-1} \prod_{j=i+1}^{n} p \left| X_{ij} \mid F(s_j - s_i) \right|, \tag{5}$$

which implies a log-likelihood of the form

$$Q = \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \log p \left| X_{ij} \mid F(s_j - s_i) \right|.$$

The log-likelihood Q is a function of $J = n + m + 1$ parameters θ_q $q = 1, \ldots, J$, these are n scale value parameters s_j, m transformation parameters a_q and the location parameter F_0.

An equality condition is needed to fix the location of the stimulus parameters. A scale constraint although in principal not strictly necessary for fixed knots is in fact very helpful for computation. For example the location may be fixed by requiring that the mean value of the stimulus parameters be zero and the scale may be fixed by fixing the maximum range of the stimulus parameters.

Inequality conditions $a_q \geq 0$ $(q = 1, \ldots, m)$ are also needed to ensure that the integral transformation remains monotone. Additional inequality constraints are needed for binary data, to ensure that $0 \leq F_{ij} \leq 1$. The number and the nature of the inequality constraints appropriate for each type of data will be further discussed in the section dealing with that specific type of data. The maximization of Q is thus a nonlinear programming problem subject to two linear equality constraints and a number of linear inequality constraints.

In the penalty function approach to solving the nonlinear programming problem with constraints, a sequence of unconstrained problems is solved. Each constraint can be expressed by requiring that a function g of the parameters to be estimated be nonnegative. To minimize $-Q$ subject to a set of M constraints, a penalty function U is chosen which is a function of the constraints $g_q(\theta_1, \ldots, \theta_J)$, $q = 1, \ldots, M$ and a penalty parameter r. One type of penalty function has the following properties:

(a) if the constraint $g_q(\theta_1, \ldots, \theta_J)$ is positive, U must tend to infinity as g_q tends to zero;
(b) for positive values of g_q U must tend to zero as r tends to infinity.

Then the values of θ_q $q = 1, \ldots, J$ that minimize the composite function

$$\Phi(\theta_1, \ldots, \theta_J, r) = -Q(\theta_1, \ldots, \theta_J) + U[g_1(\theta_1, \ldots, \theta_J), \ldots, g_M(\theta_, \ldots, \theta_J), r] \tag{6}$$

are determined using an algorithm for unconstrained optimization for each of an increasing sequence of values of r. Starting values of $g_q(\theta_1, \ldots, \theta_J)$ for $q = 1, \ldots, M$ must be positive. The particular penalty function chosen for each of the two types of data under consideration will be presented in the section dealing with that specific type of data.

Other approaches are possible and may prove superior. Among these are Newton Raphson and Scoring algorithms modified to allow for constraints. However, the generality and ease of programming the penalty function method makes it attractive, at least in early stages.

To compare the log-likelihood for the proposed integrated B-spline model for binary data against a more general model of which it is a special case one must make some decision as to how many parameters are actually being fit. At most, this would be $n + m - 1$. However, some constraint functions may take values on the boundaries. While these "tight" constraints may be viewed as reducing the number of mathematically independent parameters, it is also the case that these constraints may be considered as at least partially estimated from the data since the data determined their location at or away from the boundary. A conservative decision is to set the number of estimated parameters to

$J = m + n - 1$ when computing degrees of freedom. The asymptotic variance-covariance matrix for the parameter estimates may be computed using the Moore-Penrose inverse [Ramsay, 1978]. For a fixed value of the argument, the asymptotic variance for the transformation conditional on a fixed argument is given by

$$\text{var}\left| F(x, \boldsymbol{\theta}) \,|\, \boldsymbol{\theta} \right| = \frac{\delta F'}{\delta \boldsymbol{\theta}} \sum \frac{\delta F}{\delta \boldsymbol{\theta}} \tag{7}$$

where $\boldsymbol{\theta}$ is the vector of parameters and $\sum$ is the variance-covariance matrix of the parameter estimates. This relation can be extended to yield a variance-covariance matrix for a set of function values. It should be stressed, however, that our approach to interval estimation and hypothesis testing should be considered as giving only rough results. The usual asymptotic theory upon which it is based does not hold when parameters lie on the boundary of the parameter space [Chernoff, 1954; Feder, 1968] and is of very limited value when they lie close to it. This and better methods will undoubtedly require justification by Monte Carlo sampling, rather than faith alone.

Binary Data

For binary data X_{ij} is an observation of a random variable with a binomial probability function with parameters N and F_{ij} (assuming $F_{ij} \neq 0, 1$). When all pairs of stimuli are presented for judgment $\binom{n}{2}$ independent binomial variables are observed, and the log-likelihood excluding constant terms for the entire set of observations is

$$Q = \sum_{i-1}^{n-1} \sum_{j=i+1}^{n} \{X_{ij} \log F_{ij} + (N - X_{ij}) \log(1 - F_{ij})\}, \tag{8}$$

where N is the number of times a particular pair of stimuli with subscripts i and j is presented for comparison. The log-likelihood Q is maximized subject to the two equality conditions defining the location and scale of the stimulus parameters, and the m inequality conditions $a_q \geq 0$ ($q = 1, m$). In addition $\binom{n}{2} + 1$ inequality constraints $F_{ij} < 1$ and $F_0 > 0$ are needed to insure that $0 < P_{ij} < 1$, so the number of inequality constraints are $M = m + 1 + \binom{n}{2}$.

We chose the following penalty function to satisfy these inequality constraints:

$$U(\theta_1, \ldots, \theta_J, r) = -\frac{\log P_0}{r} - \sum_{q=1}^{m} \frac{\log a_q}{r} - \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \frac{\log(1 - F_{ij})}{10^3 r}. \tag{9}$$

In this case a reasonable test of the integrated B-spline model would be to test a completely general model

$$\Omega\colon 0 < P_{ij} < 1$$

against the special model

$$\omega\colon P_{ij} = F(s_j - s_i),\ 0 < P_{ij} < 1$$

under the alternative hypothesis that $\underline{0} \leq \underline{F}_{ij} \leq 1$. Assuming all pairs are presented for choice the likelihood ratio for testing the integrated B-spline model is

$$\lambda = \frac{\displaystyle\prod_{i=1}^{n-1} \prod_{j=i+1}^{n} \binom{N}{X_{ij}} F_{ij}^{X_{ij}} (1 - F_{ij})^{N - X_{ij}}}{\displaystyle\prod_{i=1}^{n-1} \prod_{j=i+1}^{n} \binom{N}{X_{ij}} Z_{ij}^{X_{ij}} (1 - Z_{ij})^{N - X_{ij}}}, \tag{10}$$

where Z_{ij} is a maximum likelihood estimator of the discrimination probability for stimuli i and j and it is the observed proportion of judgments j over i.

The statistic

$$Y = -2 \log \lambda = 2N \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \left[Z_{ij} \log \frac{Z_{ij}}{F_{ij}} - (1 - Z_{ij}) \log \frac{(1 - Z_{ij})}{(1 - F_{ij})} \right] \tag{11}$$

has a distribution which converges in probability with $N \rightarrow \infty$ to the chi-square distribution with degrees of freedom equal to the number of independent parameters in Ω minus the number in ω. For the integrated B-spline model a conservative choice is to set the number of independent parameters equal to $n + m - 1$. Thus the number of degrees of freedom associated with Y for testing the proposed model are $n(n - 1)/2 - (n + m - 1)$. For testing the Thurstone model or the BTL model the number of parameters is $n - 1$. Hence the appropriate number of degrees of freedom associated with Y is $(n - 1)(n - 2)/2$.

Continuous Data

For continuous data one may assert that the preference strengths X_{ij} are independently normally distributed with constant variance σ^2 and expectation F_{ij}. The log-likelihood then takes the form

$$Q = -\binom{n}{2} \log \sigma - \frac{e_{ij}^2}{2\sigma^2}, \tag{12}$$

where

$$e_{ij} = X_{ij} - F_{ij}.$$

Maximizing the likelihood in this case is equivalent to a least squares minimization. This model may be compared with the linear model. Since one model is not a specialization of the other the AIC statistic may be used to compare the two [Aikaike, 1974]. The AIC statistic is essentially the log of the likelihood L corrected for the number of parameters being estimated and is given by

$$\mathrm{AIC} = \log L - N_p$$

where N_p is the number of free parameters. It is a measure of quality of fit per parameter used in fitting. Two AIC statistics are compared in terms of magnitude, with the fit producing the larger AIC statistic being preferred. An alternative to the AIC statistic, has been proposed by Schwarz [1978]. The alternative is to choose the model that gives the largest value of

$$\mathrm{BIC} = \log L - \tfrac{1}{2} N_p \log N_0$$

where N_0 is the number of observations. Qualitatively both the BIC and AIC procedures give a mathematical formulation of the principle of parsimony in model building. The BIC procedure, which is based on the asymptotic behavior of Bayes estimators under priors which specify that the correct model is one of the models under consideration, leans more than the AIC procedure towards lower-dimensional models. Thus AIC is less conservative than BIC, and in cases of disagreement (and in any case) other considerations should be allowed to have their say. Steiger and Lind [1980, Note 3] have done some Monte Carlo comparisons of these statistics for tests of the number of common factors in factor analysis. Since the maximum likelihood estimation process for this model can be viewed as a nonlinear regression problem, one may also compute and compare the multiple correlation coefficient for each model. The multiple correlation coefficient is the correlation between the data and the fitted values on the basis of the model.

Examples

The first set of examples demonstrates the algorithm's ability to recover a set of underlying stimulus values and a transformation observed with error. A set of n stimuli with values equal to the quantiles of the standard normal distribution were chosen with $n = 7$, 10,14. These values of n double the number of observations $\binom{n}{2}$ for each subsequent value of n. The stimuli were ordered randomly and both types of data were generated for each stimuli set. For the binary data population values F_{ij} were generated using the logistic function,

$$F_{ij} = \frac{1}{1 + \exp[-(s_j - s_i)]}. \tag{13}$$

A random binomial observation X_{ij} was then generated for each F_{ij} for a sample of size $N = 100$.

For continuous data, population values F_{ij} were generated using the monotone transformation

$$F_{ij} = -\frac{1}{2} - \frac{1}{1 + \exp[-(s_j - s_i)]}. \tag{14}$$

An observation X_{ij} was then generated by adding a random normal deviate with the standard deviation 20% of the F_{ij}-value.

Order-two splines with one interior knot were used in each case. This choice was made because we found that it produces a good fit while requiring a minimal number of transformation parameters. For the first run the single interior knot was always located at zero. The exterior knots were located so as to ensure that all pairwise differences would lie within the exterior knots. Exterior knots were always located such that after the first penalty iteration, all iterated values of $(s_j - s_i)$ lay between the exterior knots. During the first iteration ($r = 1$), if $(s_j - s_i)$ lay outside the boundary the monotone transformation of $(s_j - s_i)$ was set equal to its value at the boundary. If these conditions were violated new exterior knots were chosen and the entire analysis was redone with new knots. In general it was not necessary to readjust the exterior knots, although some familiarity with the algorithm was required in order to choose values. To improve conditioning and performance of the algorithm a second run was often performed moving either the first or last two knots closer to the origin since the particular series of differences generated by the ordering of the stimuli was often not symmetric about the origin. Very occasionally if the distribution of differences about the origin was extremely assymmetric, it was desirable to move the interior knot away from the origin to improve conditioning. Moving the knots in this way had little effect on the log-likelihood in our experience, but it improved performance of the optimization procedure. It should be noted that the starting values for the stimulus-object values were randomly assigned. The results for the log-likelihood for the binary data are displayed in Table 1. Also included in the Table are the values of the Wilks test statistic Y and the knots. The results for the continuous data are displayed in Table 2. Shown in Table 2 are the log-likelihood, the squared multiple correlation coefficient, the AIC statistic and the BIC statistic for both the integrated B-spline model and the linear model.

A comparison of the input and estimated stimulus-object parameter values for the cases $n = 7$ for binary data and $n = 10$ for continuous data are presented in Figure 2. These results are fairly typical. The resultant transformations with upper and lower 95% conditional asymptotic confidence bounds and the input transformations are displayed in Figure 3. The important features of the transformation are recovered. Some improvement in the conditioning and the log-likelihood could be achieved by readjusting the knots. In our experience the improvement was minor. Part of the discrepancy between "true" and

TABLE 1

Binary data - Artificial Cases

	n = 7	n = 10	n = 14
Goodness-of-fit statistic	15.0	38.5	71.6
df	12	33	75
p	.10<p<.25	p < .10	p < .25
Average absolute discrepancy observed-derived proportions	0.031	0.030	0.030
Knots	-2.0,-2.0,0.0, 4.0,4.0	-4.5,-4.5,0.0, 6.5,6.5	-4.0,-4.0,0.0, 5.5,5.5

"fitted" transformation is due to the fact that the estimated scale values have a somewhat greater range compared with the "true" ones in the binary data case, and a somewhat smaller range in the continuous data case (see Figure 2). The "true" transformation shown in Figure 3 is the "true" transformation of the estimated difference between a pair of stimuli values. One might also compare $F(s_j - s_i)$ estimated with "true P_{ij}" for each pair of stimuli. The discrepancy between "true" and estimated values is somewhat less in this last comparison than that shown in Figure 3.

The effect of increasing the number of parameters was also investigated for $n = 10$ for the continuous data treated above. The number of parameters was increased both by increasing the number of interior knots and the order. The results are displayed in Table 3. Some improvement is obtained by adding one more parameter. The change seems to be about the same whether one adds a parameter by means of adding an interior knot or increasing the order by one.

The next example deals with two of the sets of binary data analyzed by Hohle [1966]. The first consists of preference judgments obtained from 148 observers for each pair of nine common vegetables [Guilford, 1954]. In the second set seven weights weighing from 185 to 215 grams were used as stimuli. A single subject was presented each pair of weights 100 times and was asked to judge which was heavier [Guilford, 1931]. Again order-two splines were used with one interior knot located at zero. The results for the integrated B-spline

TABLE 2

Continuous data - Artificial Cases

	n = 7		n = 10		n = 14	
	IBSM	LM	IBSM	LM	IBSM	LM
Log L	60.5	56.6	115.3	102.1	234.9	207.7
R^2	.986	.948	.974	.947	.970	.937
AIC Statistic	51.5	50.6	103.3	93.1	218.9	194.7
BIC Statistic	46.8	47.5	92.5	85.0	206.9	177.4
Knots	-3.5,-3.5,-0.5, 2.5,2.5		-2.25,-2.25,0.5, 3.5,3.5		-3.0,-3.0,0.0, 3.5,3.5	

Note: IBSM denotes integrated B-spline model
LM denotes linear model

model, the Thurstone model, and the Bradley, Terry, Luce model are presented in Table 4. None of these models offers a satisfactory fit for the vegetable data as all of the chi-square values are significant at the 0.05 level or less. The proposed model seems to fit the vegetable data better than the Thurstone model and as well as the BTL model. All of the models fit the weights data since none of the chi-square values are significant. The BTL model seems to offer the best fit. Mosteller [1958] analyzed the vegetable data using the following distributions: the uniform on the interval 0 to 1, arc sin $p^{\frac{1}{2}}$, case V (normal), the double exponential ($\frac{1}{2}$ exp $[-|x|]$), and t_{10}. Although the double exponential gave the best fit, Mosteller found that the scale values did not differ appreciably.

The third example presents continuous data for different leisure time activities. A set of ten and set of fifteen activities were presented pairwise to each of fifteen subjects. The set of ten is a subset of the set of fifteen. Order two splines were used with one interior knot. The results for each subject are presented in Table 5. Included are the squared multiple correlation coefficient R^2 for the integrated B-spline and the linear model, and the difference between the AIC and BIC statistics for the two models.

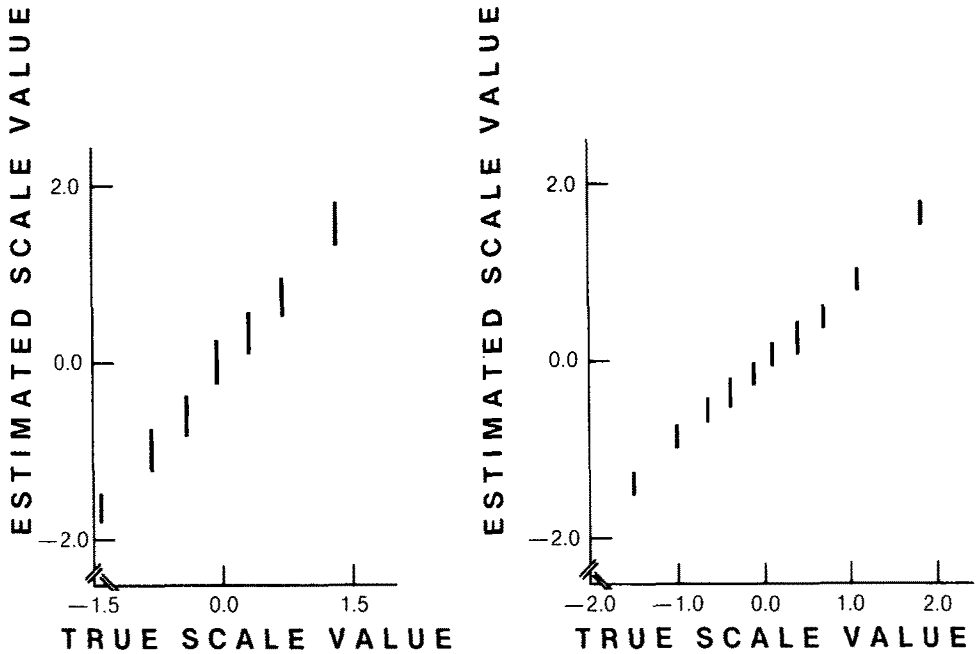

FIGURE 2
Estimated versus true stimulus values for artificial data with error
Left: $n = 7$ binary data; Right: $n = 10$ continuous data.

TABLE 3

Comparison of Integrated B-Spline Models

with Different Numbers of Parameters

Continuous Data   Artificial Case n=10

Case Number	Number of Spline Parameters	Order of B-Splines	Number of Interior Knots	Log L	A I C Statistic	B I C Statistic	R^2
1	3	2	1	117.1	105.1	94.3	.974
2	4	2	2	121.4	108.4	96.7	.978
3	4	3	1	121.0	108.0	96.3	.978
4	5	2	3	121.3	107.3	94.7	.978

Note: Case 1 Knots -2, -2, 0.5, 3.4, 3.4

Case 2 Knots -1.9, -1.9, 0.25, 1.25, 3.4, 3.4

Case 3 Knots -2, -2, -2, 0.5, 3.4, 3.4, 3.4

Case 4 Knots -1.9, -1.9, -.25, .5, 1.25, 3.4, 3.4

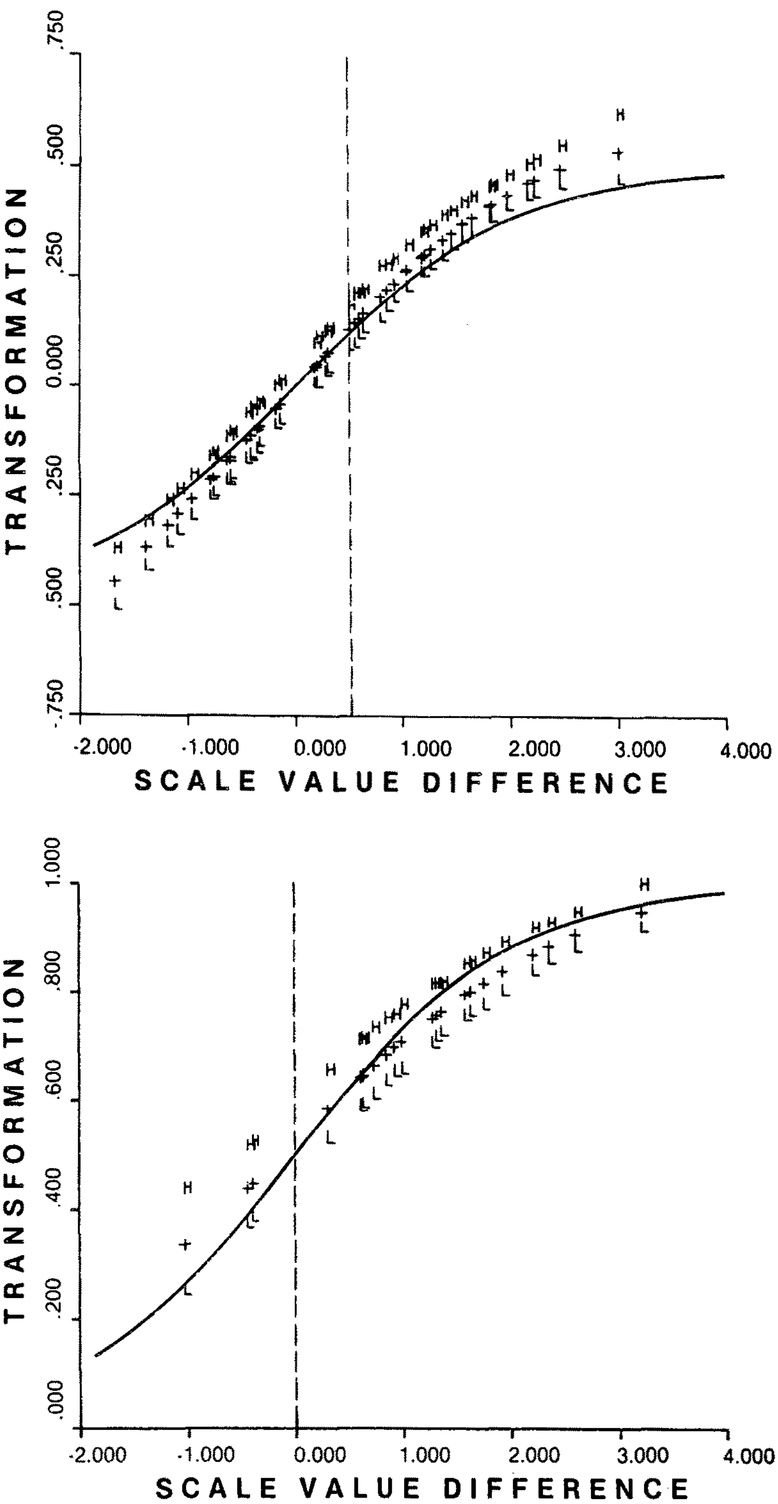

FIGURE 3

Fitted transformations for the analysis of artificial data with error. The upper graph is based on continuous data, with $n = 10$. The lower graph is based on binary data with $n = 7$. Legend: $+$, fitted transformation; H, upper bound of 95% conditional asymptotic confidence interval; L, lower bound; vertical dashed lines, position of interior knots. The solid curve is the true transformation.

TABLE 4

Goodness-of-fit Data Comparing

the Thurstone, the BTL, and the Integrated B-Spline Models

	Vegetable Data $n = 9$			Lifted Weights Data $n = 7$		
	Y	df	Error	Y	df	Error
Thurstone Model	52.91**	28	.034	10.18	15	.017
BTL Model	46.76*	28	.032	9.84	15	.017
Integrated B-Spline Model	42.95*	25	.030	12.04	12	.019

Note: ** significant at the 0.01 level

* significant at the 0.05 level

Error = average absolute discrepancy derived-observed proportions

The transformations and the utility scales obtained for two subjects are presented in Figure 4. Figure 4a presents the results for a typical subject and Figure 4b presents the results for a rather atypical subject. Upper and lower conditional asymptotic 95% confidence bounds are shown. The results indicate that in general the monotone transformation is an improvement over the linear transformation, but that the respective fits vary from subject to subject.

Discussion

One of the problems encountered in these analyses was that of locating the exterior knots. This problem arises from the transformation of a function whose domain depends on the parameter values being estimated. For continuous data this problem may be circumvented in the following way. One may assert that difference on the utility scale is monotonically related to the pairwise preference strengths. That is

$$s_j - s_i = F(X_{ij}) = \int_{x0}^{x_{ij}} v(x) + F_0. \tag{15}$$

In this case we are fitting the inverse of the transformation defined in (1). One could then assert that the transformed pairwise preference strengths are independently normally distributed with constant variance σ^2 and expectation $(s_j - s_i)$. The log-likelihood then takes the form

$$Q = -\left(\frac{n}{2}\right) \log \sigma - \frac{e_{ij}^2}{(2\sigma^2)} + \log \frac{dF(X_{ij})}{dX_{ij}}$$

TABLE 5

Comparison of Two Models Used to Analyze

Continuous Data for Leisure Activities

	n=10				n=15			
Subject	R^2(LM)	R^2(IBSM)	Δ AIC	Δ BIC	R^2(LM)	R^2(IBSM)	Δ AIC	Δ BIC
1	.529	.810	14.0	11.3	.835	.843	-1.2	-5.2
2	.588	.674	-0.4	-3.1	.806	.812	-1.9	-5.9
3	.776	.810	-1.0	-3.7	.899	.903	-3.0	-7.0
4	.780	.799	-1.7	-4.4	.929	.939	4.8	0.8
5	.792	.852	2.5	-0.2	*			
6	.817	.834	-2.8	-5.5	.861	.872	1.4	-2.6
7	.843	.865	0.1	-2.6	.823	.837	-1.7	-5.2
8	.863	.914	6.1	3.4	.894	.924	12.0	8.0
9	.867	.914	6.5	3.8	.841	.884	13.5	9.5
10	.876	.897	0.5	-2.2	.885	.931	23.7	19.7
11	.876	.885	-2.0	-4.7	.821	.846	4.8	0.8
12	.885	.941	12.0	9.3	.880	.910	11.6	7.6
13	.887	.925	5.4	2.7	.882	.906	8.0	4.0
14	.891	.920	2.7	0.0	.887	.918	12.7	8.7
15	.914	.916	-3.1	-5.8	.876	.910	13.2	9.2

Note: * No data were available for this subject for n=15

LM denotes linear model

IBSM denotes integrated B-spline model

Δ AIC = IBSM AIC statistic - LM AIC statistic

Δ BIC = IBSM BIC statistic - LM BIC statistic

where

$$e_{ij} = F(X_{ij}) - (s_j - s_i).$$

The advantage of this alternative model is that the first k knots t_q $q = 1, \ldots, k$ can be set equal to $\min\{X_{ij}\}$ and the last k knots t_q $q = m + 1, \ldots, m + k$ can be set equal to $\max\{X_{ij}\}$. However, if one is planning to analyze both continuous and binary data the use of the proposed model allows the two types of data to be treated in a similar manner.

The algorithm generally proceeded to a solution reliably. Improvements in the log-likelihood were minimal after only a few steps in the penalty parameter. Performance did depend on the choice of knots; however the likelihood and the shape of the transformation were only slightly affected by this choice. Moreover, it was usually not difficult to make a reasonable initial selection of knots.

Integrated B-splines appear to be a useful tool for representing monotone transformations. In the past in many applications ranging from the estimation of item characteristic curves to the estimation of survival curves, simple mathematical functions such as the

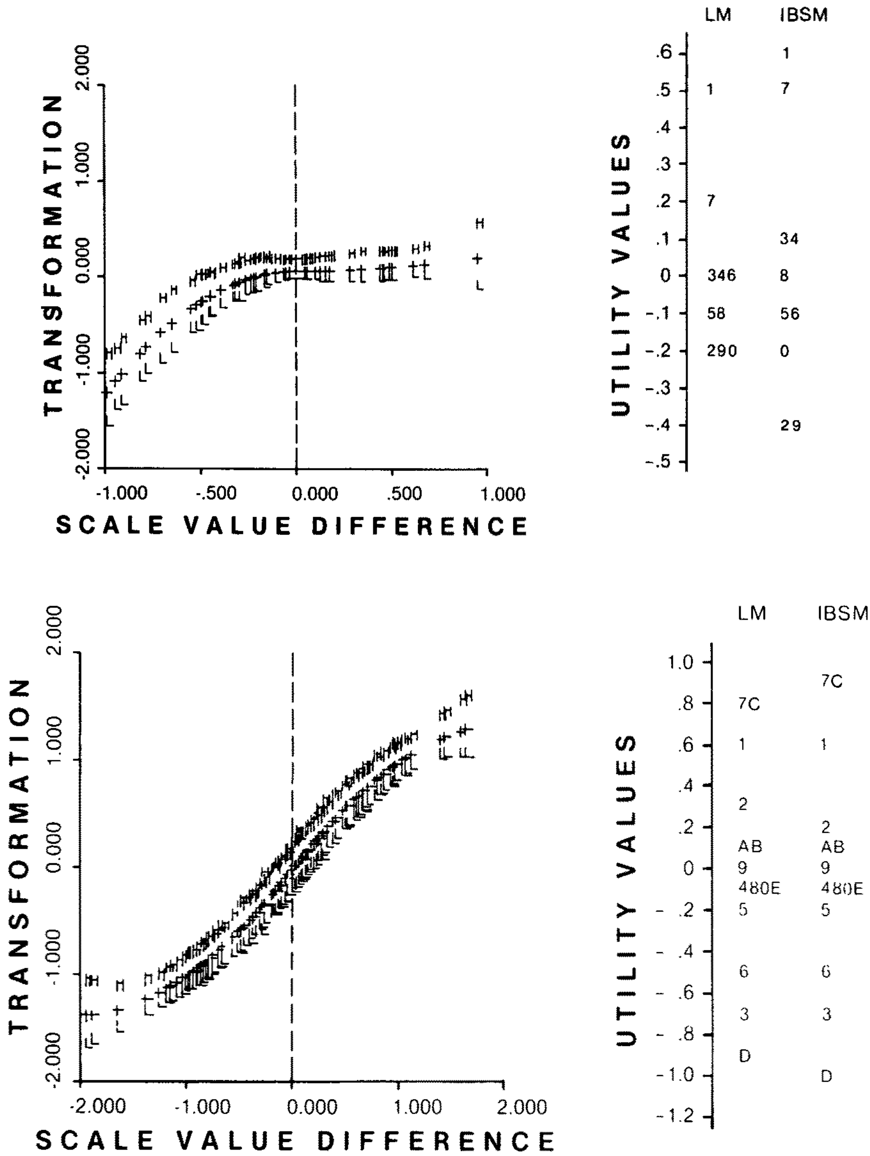

FIGURE 4

Transformations of scale values for leisure time activities data. The upper graph is for Subject #1, $n = 10$. The lower graph is for Subject #13, $n = 15$. The utility values for the stimuli are shown at the right. Legend: +, fitted transformation, H, upper bound of 95% conditional asymptotic confidence interval; L, lower bound; vertical dashed lines, position of interior knots.

logistic function or polynomials [Lord, 1968, Note 2] have been used to represent an unknown response function. For such functions their behavior in one region determines their behavior everywhere. Splines do not have this limitation. Since they are everywhere simple polynomials, they are computationally convenient, and the resulting curve is smooth. Integrated B-splines are also valuable as fixed response functions to replace the normal and the logistic functions because of their computational convenience, especially in applications like test theory where computation is slow and difficult. Finally integrated splines offer the advantage for this application that they can be constrained to have the properties of a cumulative distribution function.

REFERENCES NOTES

1. de Boor, C. & Rice, J. R. *Least squares cubic spline approximation. I–Fixed knots; II–Variable knots* (CSD TR 20 and 21). West Lafayette, Indiana: Purdue University, 1968.
2. Lord, F. *Estimating item characteristic curves without knowledge of their mathematical form* (RB-68-8). Princeton, New Jersey: Educational Testing Service, 1968.
3. Steiger, J. H. & Lind, J. M. *Statistically based tests for the number of common factors.* Paper presented at the annual meeting of the Psychometric Society, Iowa City, 1980.

REFERENCES

Akaike, H. A new look at the statistical model. *IEEE Transactions on Automatic Control*, 1974, *19*, 716–723.

Bechtel, G. *Multidimensional preference scaling.* The Hague: Mouton, 1976.

Bradley, R. A., & Terry, M. E. Rank analysis of incomplete block designs. 1. The method of paired comparisons. *Biometrika*, 1952, *39*, 324–345.

Chernoff, H. On the distribution of the likelihood ratio. *The Annals of Mathematical Statistics*, 1954, *25*, 573–578.

Cox, M. G. The numerical evaluation of B-splines. *Journal of the Institute of Mathematics and its Applications*, 1972, *10*, 134–149.

Curry, H. B. & Schoenberg, I. J. On spline distributions and their limits: The Polya distribution functions, Abstract 380 *t. Bulletin of the American Mathematical Society*, 1947, *53*, 1114.

Curry, H. B. & Schoenberg, I. J. On Polya frequency functions. IV: The fundamental spline functions and their limits. *Journal d.Analyse Mathématique*, 1966, *17*, 71–107.

de Boor, C. On calculating with B-splines. *Journal of Approximating Theory*, 1972, *6*, 50–62.

de Boor, C. *A practical guide to splines.* New York: Springer Verlag, 1978.

Feder, P. I. On the distribution of the log-likelihood ratio test statistic when the true parameter is "near" the boundaries of the hypothesis regions. *The Annals of Mathematical Statistics*, 1968, *39*, 2044–2055.

Ford, L. R. Jr. Solution of a ranking problem from binary comparisons. Herbert Ellsworth Slaught Memorial Papers. *American Mathematics Monthly*, 1957, *64*, 28–33.

Guilford, J. P. Some empirical tests of the method of paired comparisons. *Journal of General Psychology*, 1931, *5*, 64–77.

Guilford, J. P. Psychometric methods (2 ed.). New York: McGraw-Hill, 1954.

Hohle, R. H. An empirical evaluation and comparison of two models for discriminability scales. *Journal of Mathematical Psychology*, 1966, *3*, 174–183.

Kruskal, J. B. Analysis of factorial experiments by estimating monotone transformation of the data. *Journal of the Royal Statistical Society, Series* B, 1965, *27*, 251–263.

de Leeuw, J., Young, F. W., & Takane, Y. Additive structures in qualitative data: An alternating least squares method with optimal scaling features. *Psychometrika*, 1976, *41*, 471–503.

Luce, R. D. Individual choice behavior. New York: Wiley, 1959.

Mosteller, F. The mystery of the missing corpus. *Psychometrika*, 1958, *23*, 279–289.

Ramsay, J. O. Monotonic weighted power tranformations to additivity. *Psychometrika*, 1977, *42*, 83–109.

Ramsay, J. O. Confidence regions for multidimensional scaling analysis. *Psychometrika*, 1978, *43*, 145–160.

Schwarz, G. Estimating the dimension of a model. *The Annals of Statistics*, 1978, *6*, 461–464.

Takane, Y., Young, F. W. & de Leeuw, J. An individual differences additive model: An alternating least squares method with optimal scaling features. *Psychometrika*, 1980, *45*, 183–209.

Thurstone, L. L. A law of comparitive judgement. *Psychological Review*, 1927, *34*, 273–286.

Winsberg, S. & Ramsay, J. O. Monotonic Transformations to Additivity using splines. *Biometrika*, 1980, *67*, 669–674.

Manuscript received 8/15/80
Final version received 2/19/81